

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 08:44:05 ; Search time 4394 Seconds
(without alignments)
10712.457 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086

Sequence: 1 atgaataagattttatttat.....caagtagtcaaacactctag 1086

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vb.*

15: em_hum.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: sm_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	6	AX352333 Sequence
2	1086	100.0	7239	1	AF296132 Escherich
3	1086	100.0	7239	6	AX352351 Sequence
4	1086	100.0	7239	6	AX741418 Sequence
5	852.4	78.5	7174	1	ECOCFAIA
6	310.8	28.6	5336	1	ECCOCCD
7	310.8	28.6	5336	6	AX741413 Sequence
8	309.6	28.5	6956	1	AY216491 Escherich
9	298.8	27.5	5798	1	ECCOTABCD
10	298.8	27.5	5798	6	AX741415 Sequence
C 11	53.4	5.8	1141	6	AX083744 Sequence
C 12	57.6	5.3	61864	3	C3Y50B5A
C 13	53.6	4.9	5230	8	AF253047
C 14	50.6	4.7	29143	3	AC115594
C 15	50.6	4.7	88549	3	AC116979 Dictyoste
C 16	50.2	4.6	133477	2	AC006278 Plasmodiu
C 17	50.2	4.6	231551	3	AE014844
C 18	50	4.6	2000	6	AX655393 Sequence
C 19	49.6	4.6	253924	3	AE014822 Plasmodiu
C 20	49.4	4.5	10279	6	AX279996 Sequence
C 21	49.4	4.5	10279	6	AX346493 Sequence
C 22	49.4	4.5	10279	6	AX356452 Sequence
C 23	49.2	4.3	233151	3	AE014842 Plasmodiu
C 24	48.6	4.5	1141	6	AX083744 Sequence
C 25	48.2	4.4	9504	6	AX323793 Sequence
C 26	48.2	4.4	349980	6	AX344570 Sequence
C 27	48	4.4	3549	3	AF310884 Dictyoste
C 28	47.8	4.4	128683	2	AC135793 Oryza sat
C 29	47.6	4.4	169794	2	AC004688 Plasmodiu
C 30	47.6	4.4	196149	2	AC004709 Plasmodiu
C 31	47.6	4.4	252650	3	AE014847 Plasmodiu
C 32	47.2	4.3	12940	3	AE001386 Plasmodiu
C 33	46.8	4.3	105682	3	AC116957_3
C 34	46.6	4.3	110000	2	PFMAL13_09
C 35	46.6	4.3	110000	2	PFMAL13_10
C 36	46.6	4.3	132254	3	PFMAL6PL_05
C 37	46.6	4.3	132254	3	AC116330 Dictyoste
C 38	46.2	4.3	76568	3	MSREV
C 39	46.2	4.3	110000	3	AC116305_1
C 40	46	4.2	111882	3	AC115612 Dictyoste
C 41	46	4.2	153751	3	AC116551 Dictyoste
C 42	46	4.2	176436	2	EX323051 Danio rer
C 43	46	4.2	236521	2	AC129274 Rattus no
C 44	46	4.2	331039	3	AC116988 Dictyoste
C 45	45.6	4.2	251448	3	AE014819 Plasmodiu

ALIGNMENTS

RESULT 1	AX352333	1086 bp	DNA	linear	PAT 06-FEB-2002
AX352333	Sequence 9 from Patent WO0181582.				
LOCUS	AX352333				
DEFINITION	AX352333				
ACCESSION	AX352333.1				
VERSION	GI:18617616				
KEYWORDS	Escherichia coli				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
REFERENCE	1				
AUTHORS	Altboum, Z., Levine, M.M. and Barry, E.Y.				
TITLE	Isolation and characterization of the csa operon (etec-cs4 pill)				
	and methods of using same				

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 0181582-A 9 01-NOV-2001;

University of Maryland, Baltimore (US)

Location/Qualifiers

FEATURES

source

1. .1086

/organism="Escherichia coli"

/mol_type="unassigned DNA"

/db_xref="taxon:562"

1. .1086

/notes="unnamed protein product"

/codon_start=1

/translation="1"

/protein_id="CAD22854.1"

/gb_xref="GI:18617617"

/translacion="MNKILIFTFPFSVLTFVAVSADKIPGDESITNIFRPRDRNES

SPXNINNHITAVSESHITVDRMTFLCLSHNTLNACPTSENSSSVSGETNITL

QFTEKSLIARELIQYKOLLKPSVNCPSGLTNSAHFNCKNNAASGLVLYIPAG

ELKLPFGGLWDAILKLRVRSYETGTYIITIKLTDKGNLQIWLPPKPSDAVD

LAURPTGGTYIGNSVDMCFYDGYNSSSLERFQDNPNKPSDKGFYLRKINDDTKE

IAYTSLLAGKSLPTNGTSLNIAADAASLETKWNRITAVTMEISVPIVLCWPGRLQL

DAKVENPEAQYMGNIINVTTPSSQTL"

ORIGIN

Query Match 100.0%; Score 1086; DB 6; Length 1086;

Best Local Similarity 100.0%; Pred. No. 3.9e-223;

Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATAAGATTATTTATTTTACATGTTTCTCTTCAGTACATTTTACATTTGCT 60
 DB 1 ATGAATAAGATTATTTATTTTACATGTTTCTCTTCAGTACATTTTACATTTGCT 60
 QY 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCATAAATAATTTTGGCCCGGTGAC 120
 DB 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCATAAATAATTTTGGCCCGGTGAC 120
 QY 121 AGGAACGATCTCCCAACAAATATATTAATTAACATATTAAGCATAAGTGAA 180
 DB 121 AGGAACGATCTCCCAACAAATATATTAATTAACATATTAAGCATAAGTGAA 180
 QY 181 AGTCATATCTGTATGATGAGTACATTTTATATGTTGTTCTTCCCAATACACTTAAT 240
 DB 181 AGTCATATCTGTATGATGAGTACATTTTATGTTGTTCTTCCCAATACACTTAAT 240
 QY 241 GGACATGTCCAAACAGTGAAGATCTAGCAGTTCATCGTCAAGGTCGAACAAATATA 300
 DB 241 GGACATGTCCAAACAGTGAAGATCTAGCAGTTCATCGTCAAGGTCGAACAAATATA 300
 QY 301 ACATTACATTTACGGAAGAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
 DB 301 ACATTACATTTACGGAAGAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTAT 360
 QY 361 AAACAATTTATTTCAAAAGTGTAACTGCGCCATCCGCGCTTAACACTTAACACTCAT 420
 DB 361 AAACAATTTATTTCAAAAGTGTAACTGCGCCATCCGCGCTTAACACTTAACACTCAT 420
 QY 421 TTTAACTGTATAAAGACCGCGCTTCAAGTGAAGTTTATATATTTATATTCCTGCTGGC 480
 DB 421 TTTAACTGTATAAAGACCGCGCTTCAAGTGAAGTTTATATATTTATATTCCTGCTGGC 480
 QY 481 GAACATAAAATTTGCTTTTGGTGTATCTGGATGCTACTCTCAAGTTAAGAGTAAAA 540
 DB 481 GAACATAAAATTTGCTTTTGGTGTATCTGGATGCTACTCTCAAGTTAAGAGTAAAA 540
 QY 541 AGACGATATAGTGAACCTATGGAACCTTACACTATAAATATCACTTATTAATTAACAT 600
 DB 541 AGACGATATAGTGAACCTATGGAACCTTACACTATAAATATCACTTATTAATTAACAT 600
 QY 601 AAGGAATATTCAGATATGTTACTCACTTCAAAAGTCAAGTCCGGTCACTTAAC 660
 DB 601 AAGGAATATTCAGATATGTTACTCACTTCAAAAGTCAAGTCCGGTCACTTAAC 660
 QY 661 TTGGCTCAACTGTGGGGGCACATATATTTGGAAGAAATCTGTTGATATGCTTTTAT 720

Db 661 TTGGCTCAACTGTGGGGGCACATATATTTGGAAGAAATCTGTTGATATGCTTTTAT 720
 QY 721 GATGGATATAGTACTAATACAGACGCTCTTTGGAGATAAGATTTTCAGGATACAAATCCTAAA 780
 Db 721 GATGGATATAGTACTAATACAGACGCTCTTTGGAGATAAGATTTTCAGGATACAAATCCTAAA 780
 QY 781 TCTGATGGGAATTTTATCTAAGGAAAAATTAATGATGACACCAAGAAATGATATACT 840
 Db 781 TCTGATGGGAATTTTATCTAAGGAAAAATTAATGATGACACCAAGAAATGATATACT 840
 QY 841 TTCTCACTCTCTTGGGGGTAAAAGTTTAACTCCAAATGGAACCTCATTAATATTT 900
 Db 841 TTCTCACTCTCTTGGGGGTAAAAGTTTAACTCCAAATGGAACCTCATTAATATTT 900
 QY 901 GCTGACGACGCTTCTCTGGAACAAACTGGAATAGAAATACAGCTGTACACATGCCAGAA 960
 Db 901 GCTGACGACGCTTCTCTGGAACAAACTGGAATAGAAATACAGCTGTACACATGCCAGAA 960
 QY 961 ATCAGTGTTCGGTGTGTTTGGCCCTGGACGTTTGAATGATGCAAAAGTGGAAAT 1020
 Db 961 ATCAGTGTTCGGTGTGTTTGGCCCTGGACGTTTGAATGATGCAAAAGTGGAAAT 1020
 QY 1021 CCGGAGCTGCAATATATGCGGTAAATTAATTAATGTTACTTTCACCAAGTAGTCAAAACA 1080
 Db 1021 CCGGAGCTGCAATATATGCGGTAAATTAATTAATGTTACTTTCACCAAGTAGTCAAAACA 1080
 QY 1081 CTCTAG 1086
 Db 1081 CTCTAG 1086
 RESULT 2
 LOCUS AF296132 7239 bp DNA linear BCT 27-FEB-2003
 DEFINITION Escherichia coli csa operon, complete sequence, and insertion
 sequence ISI.
 ACCESSION AP296132
 VERSION AP296132.1 GI:15419711
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1. (bases 1 to 7239)
 AUTHORS Altboum, Z., Levine, M.M., Galen, J.E. and Barry, E.M.
 TITLE Genetic Characterization and Immunogenicity of Coli Surface Antigen
 4 from Enterotoxigenic Escherichia coli When It Is Expressed in a
 Shigella Live-Vector Strain
 JOURNAL Infect. Immun. 71 (3), 1352-1360 (2003)
 MEDLINE 22483692
 PUBMED 12595452
 REFERENCE 2. (bases 1 to 7239)
 AUTHORS Altboum, Z.D., Levine, M.M., Galen, J.E. and Barry, E.M.
 TITLE Isolation and characterization of ETEC CS4 fimbriae encoding genes,
 and their expression in Shigella flexneri 2a guABA strain CVD 1204
 JOURNAL Unpublished
 REFERENCE 3. (bases 1 to 7239)
 AUTHORS Altboum, Z.D., Levine, M.M. and Barry, E.M.
 TITLE Direct Submission
 JOURNAL Submitted (14-AUG-2000) Center for Vaccine Development, University
 of Maryland, School of Medicine, 685 W. Baltimore Street,
 Baltimore, MD 21201, USA
 FEATURES
 source
 1. .7239
 /organism="Escherichia coli"
 /mol_type="Genomic DNA"
 /strain="E11881A"
 /serotype="O25:H42"
 /db_xref="taxon:562"
 /note="enterotoxigenic"
 283. .6849
 /gene="csa operon"
 /note="encodes ETEC CS4 fimbrial proteins; CFA/IV"


```

Qy 661 TTGGCTCCAACTGTGGGGGACATATATTTGGAAGAAATCTGTGTGATATGTGCTTTAT 720
Db 4856 TTGGCTCCAACTGTGGGGGACATATATTTGGAAGAAATCTGTGTGATATGTGCTTTAT 4915
Qy 721 GATGGATATAGTACTAAACAGCAGCTCTTGGAGATAGATTTCAGGATAACATCCCTAAA 780
Db 4916 GATGGATATAGTACTAAACAGCAGCTCTTGGAGATAGATTTCAGGATAACATCCCTAAA 4975
Qy 781 TCTGATGGAAATTTTATCTAAGGAAATAAATGATGACACAAAGAAATTCGATATACT 840
Db 4976 TCTGATGGAAATTTTATCTAAGGAAATAAATGATGACACAAAGAAATTCGATATACT 5035
Qy 841 TTGTCACTTCTTGTGGGGGTAAGTTTAACTTCAACAAATGAAAGTCAATTAATTT 900
Db 5036 TTGTCACTTCTTGTGGGGGTAAGTTTAACTTCAACAAATGAAAGTCAATTAATTT 5095
Qy 901 GCTCAGCAGCTTCTCTGGAACAACTGGAATAGATTACAGCTGCACATGCCAGAA 960
Db 5096 GCTCAGCAGCTTCTCTGGAACAACTGGAATAGATTACAGCTGCACATGCCAGAA 5155
Qy 961 ATCAGTGTCCCGGTGTGTGTGGCCCTGGAGCTTTGCAATTTGGATGCAAAAAGTGGAAAAT 1020
Db 5156 ATCAGTGTCCCGGTGTGTGTGGCCCTGGAGCTTTGCAATTTGGATGCAAAAAGTGGAAAAT 5215
Qy 1021 CCCGAGGCTGACAAATATATGCGTAATTAATTAATGTTACTTTCAACCAAGTAGTCAACA 1080
Db 5216 CCCGAGGCTGACAAATATATGCGTAATTAATTAATGTTACTTTCAACCAAGTAGTCAACA 5275
Qy 1081 CTCTAG 1086
Db 5276 CTCTAG 5281

RESULT 3
AX352351 LOCUS AX352351 7239 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 27 from Patent WO0181582.
ACCESSION AX352351
VERSION AX352351.1 GI:18617634
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS Altboun, Z., Levine, M.M. and Barry, E.M.
TITLE Isolation and characterization of the cea operon (etec-cs4 pili)
JOURNAL and methods of using same
Patent: WO 0181582-A 27 01-NOV-2001;
University of Maryland, Baltimore (US)
FEATURES
Location/Qualifiers
source 1..7239
organism="Escherichia coli"
mol_type="unassigned DNA"
db_xref="taxon:562"
283..999
note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22859.1"
/db_xref="GI:18617635"
/db_xref="REMRXBL:CAD22859"
/translation="MHKLFCLLSLITFPFVANFMFIYPIKSLKNGNSLIRVYSK
KEIOYKIYTKKINPGTSEHVMNPDGFFVTPKVLIPAGGSKSLRITQFIP
KKESTIRVYFPAVDPDSKENVDKLSTELSVNLIYAALIRSPSEONSLNRSNA
RNHIIYNGNVRAGVDIYFCKSSNIDSDCKVTHKNIIYPERKSPFLVNNNPSYF
IKLNHDEIKSGGLIQKVP"
1028..1531
note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22860.1"

```

```

/db_xref="GI:18617636"
/translation="REMRXBL:CAD22860"
/translation="NKLKTIIGAMALTTMFVAMSASAVEKNITVTASVDFDILQAD
GSLTFAVELTYSFAASRFENYKATKHCNIVKLVKLVDPKLTVLDSKQLP
ITVSYGKTLSTADVTPEPAELNFTSGVTGVSSQDLVIGATTAAQAPTAGNYGVVS
ILMTLAS"
1589..4192
note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22861.1"
/db_xref="GI:18617637"
/db_xref="REMRXBL:CAD22861"
/translation="MTKNLYITIIAMLTIPKVSFGDIPNSFRDLWGEQDEYEVKL
YQGLGHIKKTPTTHIKFYPESILDKNPKXKKEKKJSLFTNSFRNGNMSCQN
ATTQNCNVIKTSVDVI VDDVDNVNKLFIGNEFLDSEAHNDYHQLSRNVKAFQS
QTNVSDNYSKLSVSGNSALGIDTTSVAVLNMWNYKNGYNNERTINSLYFPH
DLKRYOFGMRDRTDLSQSSGNENFNLPLPDIDGTFGTQSVIKNTKDFIAPB
VTWLNFRSVEAPRANQLLGVYLDSCVNELTARLPYGSYDLKLPENTOLVREE
IIPFNKRSISIGQMDVFIQGGNIINDKRYIEKQNNHKSVMAGRLRFTTKNSVQ
QGASVDNKNYEGSLKNSGILSLSEFSLWGDNAKNTQSYISYDGFSLSFYH
NDRKVDNCGNVMAGSGCYESASLSIPLLGWSTLGYSDTYSYVKNHILSEYG
FYNONYKGRTORWOLTSLSLKWMDYNFPAIGTYSNORQJLTDKGGYTSVLTFRAS
RENLAAGYSVNYSRGKYSNELFVDGYMTSTNGDYHEVMRPNKRNHAEGRLSGR
INRPFGLNGSPMNKRNNTSNHSLTGCYNSPALTSDFYFGSGASGLTKLAGGI
IKVSNDRKKNLVKVTGALYDYSLGSNDNAFIPVPALTPASLIEBNNTGDKNLSVL
APTNDDHILFPGNVYVELBTYKVSYSYGRGFDKNGTFLSGAHVLPNEPHVLDEDDGF
SPFYTGNEKTLFLLKGRITTYTCQCGKNKHKGIVFVGDVICYNS"SSLDFEFKNPR
VQDLAKNDKG"
4196..5281
note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22862.1"
/db_xref="GI:18617638"
/db_xref="REMRXBL:CAD22862"
/translation="MHKLIPIFLFPSSVLFTFAVSADKIPGDESIINIFGRPRDMS
SPKNNHNNHTAYSESHTYDRMTFLCLSHNTLNGACCTSENPSSSSVSGENITL
QFTEKSLIKREIQIKGYKOLLKPSVNCPSGLTLNSAHFNKNAAAGSLYLIPAG
ELNLPFGGFWDATLKLVRKRYSETYGTINITIKLTDKNGIQIWLPOFKSDARVD
LNLRTPGGTYIGRNSVDMCFYDGYSTSSSLEIRFQDNNPKSDGKFLKINDTKE
IAYTSLLLAGKSLTPINGTSLNIAAASLETNNRITAVTMPBISVPLVCWPGRQLQ
DAKVENPEACQYMGNIINVTTPSSQT"
Query Match 100.0%; Score 1086; DB 6; Length 7239;
Best Local Similarity 100.0%; Pred. No. 3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGATAAGATTTTATTTATTTTACATTTGTTTCTCTTCAGTACTTTTACATTTGCT 60
Db 4196 ATGATAAGATTTTATTTATTTTACATTTGTTTCTCTTCAGTACTTTTACATTTGCT 4255
Qy 61 GTATCGGCAGATAAATTCGCGGAGATGAAGAGATAACTAATATTTTGGCCCGCTGAC 120
Db 4256 GTATCGGCAGATAAATTCGCGGAGATGAAGAGATAACTAATATTTTGGCCCGCTGAC 4315
Qy 121 AGGAACGAATCTTCCCAACATATAATATTAATAACCATATTACAGATACAGTGAA 180
Db 4316 AGGAACGAATCTTCCCAACATATAATATTAATAACCATATTACAGATACAGTGAA 4375
Qy 181 AGTCATCTCTGTATGATAGTAGACTTTTATGTTTGTCTTCTCACAATACACTTAAT 240
Db 4376 AGTCATCTCTGTATGATAGTAGACTTTTATGTTTGTCTTCTCACAATACACTTAAT 4435
Qy 241 GGAGCATGTCCACCACTGAGATCCCTAGCAGTCTATCGGTGAGCGGTGAAACAAATATA 300
Db 4436 GGAGCATGTCCACCACTGAGATCCCTAGCAGTCTATCGGTGAGCGGTGAAACAAATATA 4495
Qy 301 ACATTACAATTTACGGAAGAAAGAAAGTTTAAATAAAGACAGCTTACAAATTAAGGCTAT 360
Db 4496 ACATTACAATTTACGGAAGAAAGAAAGTTTAAATAAAGACAGCTTACAAATTAAGGCTAT 4555

```



```
QY 361 AAACRATTATTGTTCAAAAGTGTTAACTGCCATCGGCGCTAACACTTAACCTCAGCTCAT 420
DB |||||
QY 4556 AAACAATTATTGTTCAAAAAGTGTTAACTGCCATCGGCGCTAACACTTAACCTCAGCTCAT 4615
DB |||||
QY 421 TTTAACCTGTAATAAAAACGGCGCTTCAGGTCGAAGTTTATTTATATATATCTCTGCTGGC 480
DB |||||
QY 4616 TTTAACCTGTAATAAAAACGGCGCTTCAGGTCGAAGTTTATTTATATATATCTCTGCTGGC 4675
DB |||||
QY 481 GAACTAAAAAATTTGCTTTTGTGTTGATCTGGATCTGCTACTCTGAAGTTAAAGTAAAA 540
DB |||||
QY 541 AGACGATATAGTCAGACCTATGCACTTACACTATAAATCACTATTAAATTAACCTGAT 600
DB |||||
QY 4736 AGACGATATAGTCAGACCTATGCACTTACACTATAAATCACTATTAAATTAACCTGAT 4795
DB |||||
QY 601 AAGGGAATATTCAGATATATGTTTACCTCAGTTCAAAAGTCAGCTCGCGTCGATCTTAAC 660
DB |||||
QY 4796 AAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTCAGCTCGCGTCGATCTTAAC 4855
DB |||||
QY 661 TTGCGTCCAACTCGTCGGGGCACAATATATTGGAAGAAATCTGTTGATATGTCCTTTAT 720
DB |||||
QY 4856 TTGCGTCCAACTCGTCGGGGCACAATATATTGGAAGAAATCTGTTGATATGTCCTTTAT 4915
DB |||||
QY 721 GATGGATATAGTAAACAGCAGCTCTTTGGAGATAAGATTCAGGATTAACATCTTAA 780
DB |||||
QY 4916 GATGGATATAGTAAACAGCAGCTCTTTGGAGATAAGATTCAGGATTAACATCTTAA 4975
DB |||||
QY 781 TCTGATGGAAATTTTATCTAAGGAAAAATAAATGATGACACCAAGAAATTTGCATATACT 840
DB |||||
QY 4976 TCTGATGGAAATTTTATCTAAGGAAAAATAAATGATGACACCAAGAAATTTGCATATACT 5035
DB |||||
QY 841 TTGTCACCTCTCTTGGCGGTAAGAAATTTAACTCCAAAGTAAAGTAAAGTAAATTT 900
DB |||||
QY 5036 TTGTCACCTCTCTTGGCGGTAAGAAATTTAACTCCAAAGTAAAGTAAAGTAAATTT 5095
DB |||||
QY 901 GCTGACGCGACCTCTCTGGAACCAAACTGAATGATTAACAGTACAGCTGCACCATGCCAGAA 960
DB |||||
QY 5096 GCTGACGCGACCTCTCTGGAACCAAACTGAATGATTAACAGTACAGCTGCACCATGCCAGAA 5155
DB |||||
QY 961 ATCAGTGTTCGGGTGTTGTTGCGCTTGGAGTTCGAAATGGATGCAAAAGTGGAAAT 1020
DB |||||
QY 5156 ATCAGTGTTCGGGTGTTGTTGCGCTTGGAGTTCGAAATGGATGCAAAAGTGGAAAT 5215
DB |||||
QY 1021 CCGAGCGTGGACAAATATATGCGTAATATTAATGTTACTTTCACACCAAGTACTGAAACA 1080
DB |||||
QY 5216 CCGAGCGTGGACAAATATATGCGTAATATTAATGTTACTTTCACACCAAGTACTGAAACA 5275
DB |||||
QY 1081 CTCCTAG 1086
DB |||||
QY 5276 CTCCTAG 5281
DB |||||

RESULT 4
AX741418
LOCUS AX741418 7239 bp DNA linear PAT 29-MAY-2003
DEFINITION Sequence 8 from Patent WO03022306.
ACCESSION AX741418
VERSION AX74-418.1 GI:30524215
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Turner,A.K., Greenwood,J., Stephens,J.C., Beavis,J.C. and
Darsley,M.J.
Bacterial vaccine
Patent: WO 03022306-A 8 20-MAR-2003;
JOURNAL Acambis Research Limited (GB)
FEATURES
source Location/Qualifiers
1..7239
/organism="Escherichia coli"
```

```
/mol_type="genomic DNA"
/db_xref="taxon:562"
283..999
/gene="csaA"
283..999
/gene="csaA"
/codon_start=1
/transl_table=11
/protein_id="CAD90914.1"
/db_xref="GI:30524216"
/db_xref="REMRMBL:CAD90914"
/translation="MHKLPCLISLLTPFVANANFMIYPIKDLKNGNSEIIFVYKSK
KSIQIKYIKKILNPCTTEHEVDPNDGGFVTPQKVLPGKSGSRIRLTQFLP
KKEIYIRVFEAVKPSKENVINDKKLTTELSNIIYAAIRSLPSEQISLISRNA
RKNLIYNNGNRVAGVDIYFCKSSNIDDSCKVKTKHKNIIYPEKSPDLTVNNPFSYF
IKLNFEDIBK3QGLQLKVP"
1028..1531
/gene="csaB"
1028..1531
/gene="csaB"
/codon_start=1
/transl_table=11
/protein_id="CAD90915.1"
/db_xref="GI:30524217"
/db_xref="REMRMBL:CAD90915"
/translation="MKLKTIGAMALTTFVAMSASAVEKNIIVTASVDEPTIDILQAD
GSLPTAVELTSPAAAREFNKIATPKHNVINKNVLKLVNDPKLTNVLDSKQLP
ITVSGKTLTSTADVTFEPAELNFGTSGVTGVSSQDLVIGATTQAPTAGNYSYVS
ILMTLAS"
1589..4192
/gene="csaC"
1589..4192
/gene="csaC"
/codon_start=1
/transl_table=11
/protein_id="CAD90916.1"
/db_xref="GI:30524218"
/db_xref="REMRMBL:CAD90916"
/translation="MTKKNLTITII-AMLTPTYSVFGDIPNSFRDLWGODEFYEVKL
YQGTGLHRIKTPTHIKFYSPESILDKINVKKEKLSVLPFTNSRNGNSCOON
ATIQVNCNIVKTSVDVIVDDVNVNVLFIGNEFLDSEAHDEYHQLSRNKKAFIQS
QIYNVDSKYSLSVSGNSALGIDTTSYAVLNWNVKNKNGYNNERTISNLYPRH
NKRVDNCRNYNAGSGCYESASLIPLLGWTSTLGYSDTYSVVKHNLISFYH
FYNQIYKGTORTWOLTSSTLKMNDYNMPLGIGIYNSOROLTDAGGYISVLTLRAS
RENSLNAGYSYNSRKYSSNELFVDGYMTSTNNGYHEVRMPNKRHNAEGELSGR
INNRFGDNGSPMNKRNRNTSSNHSITGGYNSFALTSCGFYWGSSAGLTKLGGI
IKVKNDRTKNLVKVTGALYDYSLSGSDNAFIPVPAITPASLIIISDNNYGDKNISVL
AFTNNDMFLPGNVPEIETKYSYVIGKFDKNGTPLSGAHVNEPHVILDEDDGF
SFEYTGNEKTLFLKGRITTYTQLGKKNVHKGIYFVGVVICDVSNTSSLPDEFVKPR
VODLLAKNDKG"
4196..5281
/gene="csaE"
4196..5281
/gene="csaE"
/codon_start=1
/transl_table=11
/protein_id="CAD90917.1"
/db_xref="GI:30524219"
/db_xref="REMRMBL:CAD90917"
/translation="MNKILIFTLFFSSVYFTFAVSADKIPGDESITNIFGPRDRNES
SPKHNILNHHITQYSSSHTLYDRMTPLCLSSHTNLNCACTENPSSSVSGSETNITL
QPTXESLILKRELQIKGYKQLLPKSVNCPGSLTNSAHFNCKNKNASAGSLIYI1PAG
ELKNLPFGGIDATIKLRVKRYSYFYTYTITIKLTDKGNIQIWLQFQSDAARD
LNLPRPTGGTYICRNSVDCFYGYSTNSSSLEIRFQDNNPKSDGAFKINDDDTKE
IATYLSLLIAGKSLPTNGTCSLNIADAASLEINWNITATWPTMPEISVPLCWPLQOL
DARVNPAGQYMGNIINVTTFSSQL"
6346..6949
/feature="unlabeled protein product; IS1 sequence"
/codon_start=1
```

```

/transl_table=11
/protein_id="CAD90918.1"
/db_xref="GI:30524220"
/db_xref="RENTREMBL:CAD90918"
/translation="MFQNSPHYGRWFOHDFTSKLKLRPOSVTSRIOPSGSDVIVCBMD
EOMGYVYGAQRQWLFYADRLEKTVVAHVFGRTWATLGRMLSLSPDVIWMTDG
WPLVESRLKGLHVISRYVTQRIERHNLRLQRLARLGRKSUSFSKVELHDKVIGHY
LNKHVQ"
ORIGIN
Query Match 100.0%; Score 1086; DB 6; Length 7239;
Best Local Similarity 100.0%; Pred. NC. 3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAATAAGATTTTATTATTATTATTACATGTTTTCCTTCAGTACTTTTTTACATTTGCT 60
DB 4196 ATGAATAAGATTTTATTATTATTATTACATGTTTTCCTTCAGTACTTTTTTACATTTGCT 4255
QY 61 GTATCGGCAGATAAAATCCCGGAGATGAAGCATAACTAATATTTTGGCCCGCGTGAC 120
DB 4256 GTATCGGCAGATAAAATCCCGGAGATGAAGCATAACTAATATTTTGGCCCGCGTGAC 4315
QY 121 AGGACCAATCTCCCAACACATAATATTAATTAACCATATTACAGCATACAGTGAA 180
DB 4316 AGGACCAATCTCCCAACACATAATATTAATTAACCATATTACAGCATACAGTGAA 4375
QY 161 AGTCATACTCTGTATGATAGGATGACATTTTATGTTTGTCTTCTCAATACACTTAAT 240
DB 4376 AGTCATACTCTGTATGATAGGATGACATTTTATGTTTGTCTTCTCAATACACTTAAT 4435
QY 241 GGAGCATGTCCACCATGTAGAACTCTAGCAGTTTCNTCGGTACGGGTGAACCAATATA 300
DB 4436 GGAGCATGTCCACCATGTAGAACTCTAGCAGTTTCNTCGGTACGGGTGAACCAATATA 4495
QY 301 ACATTAATTTACGGAARAGAGTTTATATAAAGAGAGCTACAAATTTAAGGCTAT 360
DB 4496 ACATTAATTTACGGAARAGAGTTTATATAAAGAGAGCTACAAATTTAAGGCTAT 4555
QY 361 AAACAATTAATGTTCAAAAGTGTAACTGCGCATCCGCGCTTAACACTTAACCTACGCTCAT 420
DB 4556 AAACAATTAATGTTCAAAAGTGTAACTGCGCATCCGCGCTTAACACTTAACCTACGCTCAT 4615
QY 421 TTAACTGTAATAAAAAACGGCGCTCAGGTGCAAGTTTATATTTATTTATTCCTGCTGGC 480
DB 4616 TTAACTGTAATAAAAAACGGCGCTCAGGTGCAAGTTTATATTTATTTATTCCTGCTGGC 4675
QY 481 GAATCTAAAAATTTGCCCTTTTGTGATCTCGGATGCTACTCTGGAAGTTAAGAGTAAAA 540
DB 4676 GAATCTAAAAATTTGCCCTTTTGTGATCTCGGATGCTACTCTGGAAGTTAAGAGTAAAA 4735
QY 541 AGAGCATATAGTACGACCTATGGAACCTTACACTATAAATACTCACTATTAAATTAACCTGAT 600
DB 4736 AGAGCATATAGTACGACCTATGGAACCTTACACTATAAATACTCACTATTAAATTAACCTGAT 4795
QY 601 AAGGCAATATTCAGATATGTTTACCTCAGTTCAAAGTGAAGTCCGCTGCATCTTAAC 660
DB 4796 AAGGCAATATTCAGATATGTTTACCTCAGTTCAAAGTGAAGTCCGCTGCATCTTAAC 4855
QY 661 TTGGTCAACTGCTGGGGGCACATATATTTCGAGAAATTTCTGTTGATATGCTTTTAT 720
DB 4856 TTGGTCAACTGCTGGGGGCACATATATTTCGAGAAATTTCTGTTGATATGCTTTTAT 4915
QY 721 GATGATATAGTACTTAACAGACGCTCTTTGGAGATAAGATTTTCAGATAACAATCTAAA 780
DB 4916 GATGATATAGTACTTAACAGACGCTCTTTGGAGATAAGATTTTCAGATAACAATCTAAA 4975
QY 781 TCTGATGGAATTTTATCTAAGGAAATAAATGATGACCAAGAAATTCATATCTACT 840
DB 4976 TCTGATGGAATTTTATCTAAGGAAATAAATGATGACCAAGAAATTCATATCTACT 5035
QY 841 TTGTCACATCTCTCGCGGGTAAAGTTTAACTCCAAAGATGGAACGCTCATTAATATT 900
DB 5036 TTGTCACATCTCTCGCGGGTAAAGTTTAACTCCAAAGATGGAACGCTCATTAATATT 5095
```

```

QY 901 GTTGACGACGCTTCTCTGGAACAAACTGGAATAGAAATTACAGCTGTCCACATGCCAGAA 960
DB 5096 GTTGACGACGCTTCTCTGGAACAAACTGGAATAGAAATTACAGCTGTCCACATGCCAGAA 5155
QY 961 ATCAGTGTTCGGGTGTGTGTGGCTGACAGCTTTCGAATTGATGCAAAAGTGGAAAT 1020
DB 5156 ATCAGTGTTCGGGTGTGTGTGGCTGACAGCTTTCGAATTGATGCAAAAGTGGAAAT 5215
QY 1021 CCGAGCGCTGGACAATATATATGTTTAAATTTACTTTTCAACAAAGTACTCAAAACA 1080
DB 5216 CCGAGCGCTGGACAATATATATGTTTAAATTTACTTTTCAACAAAGTACTCAAAACA 5275
QY 1081 CTCTAG 1086
DB 5276 CTCTAG 5281
RESULT 5
ECOCFAIA 7174 bp DNA linear BCT 06-VAR-1995
LOCUS E.coli CFA/I fimbrial operon: cfa a, cfa b, cfa c, cfa e, cfa d'
DEFINITION genes, complete cds.
ACCESSION M55661.1 GI:145507
VERSION cfa a gene; cfa b gene; cfa c gene; cfa d' gene; cfa e gene;
KEYWORDS colonisation factor antigen I; regulatory protein.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 7174)
AUTHORS Hamers,A.M., Pel,H.J., Willshaw,G.A., Kusters,J.G., van der
Zeijst,B.A., and Gastra,W.
TITLE The nucleotide sequence of the first two genes of the CFA/I
fimbrial operon of human enterotoxigenic Escherichia coli
JOURNAL Microb. Pathog. 6 (4), 297-309 (1989)
MEDLINE 89330163
PUBMED 2569152
REFERENCE 2 (sites)
AUTHORS Gastra,W., Jordi,B.J.A.M., Mul,E.M.A., Hamers,A.M.,
McConnell,M.M., Willshaw,G.A., Smith,H.R., and van der Zeijst,B.A.M.
TITLE A silent regulatory gene cfaI on region 1 of the CFA/I plasmid NTP
113 of enterotoxigenic Escherichia coli
JOURNAL Unpublished (1990)
REFERENCE 3 (sites)
AUTHORS Jordi,B.J., Willshaw,G.A., van der Zeijst,B.A., and Gastra,W.
TITLE The complete nucleotide sequence of region 1 of the CFA/I fimbrial
operon of human enterotoxigenic Escherichia coli
JOURNAL DNA Seq. 2 (4), 257-263 (1992)
MEDLINE 92329981
PUBMED 1352712
COMMENT Original isolate E7473) enterotoxigenic DNA.
FEATURES
source
1. 7174
Location/Qualifiers
/organism="Escherichia coli"
/mol_type="genomic DNA"
/isolate="E7473"
/db_xref="taxon:562"
/clone="NTP 513"
/dev_stage="enterotoxigenic"
/gene="cfa a"
/codon_start=1
/trans_table=11
/product="colonisation factor antigen a"
/protein_id="AAC4141.1"
/db_xref="GI:145508"
/translaton="MHKLFYLLSLMAPPVANFMVPTSKLKGNSSELVRYVSKS
KEIQYIKYIKINPCTSEYKVDIPNWDGGVNVTPQKILPAGASKSIELTQFKIP

```

```

KKEEVYVYFEAVKQDSKENVNDKKLTTELSTVNIYVAAALISLPSEQNLISLNSENA
KNIIILYNNGNVRAGVKDIYFCKSSNLDNCVKANKNIYEKSPDTLWNNFSTVYF
IKLNHEGIEKQGLIQKVP"
840. .904
/gene="cfa a"
mat_peptide
905. .1561
/gene="cfa a"
/products="colonisation factor antigen a"
1593. .2105
/gene="cfa b"
CDS
1593. .2105
/gene="cfa b"
/codon_start=1
/transl_table=11
/product="colonisation factor antigen b"
/protein_id="AAC41415.1"
/db_xref="GI:145509"
/translations="MKPKTIGAMALTTFVAVSASAVEKNIIVTASVDPADLLQAD
GNALPSAVKLAYSPAKSTPESRVMTQVHTNDATKKVIVKLDAPDTQLTDLVNSTVOMP
ISVSGGQVLSITTAKEFEAAALGYSAGVNGVSSQELVISAAPKTAGTAPTAGNYSIG
VSLNLTGS"
1593. .1661
/gene="cfa b"
mat_peptide
1662. .2102
/gene="cfa b"
/products="colonisation factor antigen b"
2170. .4779
/gene="cfa c"
CDS
2170. .4779
/gene="cfa c"
/codon_start=1
/transl_table=11
/product="colonisation factor antigen c"
/protein_id="AAC41416.1"
/db_xref="GI:145510"
/translations="MKKKKRVAVALSVALIPYIGVTGDI PDSFRLWGEODEFYEV
KLYGQTLGHERIKTTHTKFYSPESILQKLNKKEKELSVFPTNSPSRNGMSQ
GNTT-QYNCNYIKTKVDVIVDDVNVN.FIGNEFLDSEAHNDEHQHSRVVKAFI
QSDTINLSGKGRKLSISGNSALGTDTSYAVLNWMNKNKSGNSNEKTINSUYF
RHLDKRYQYQFMRPTDLSQISGSFNFLLPLPIDGIRGTQTSYIKNTDKPIA
SPVTMLNPSRFAERNDOLLGWYLDGVBNDLTARIPYGSYDLKLI FENTQIVR
EZIPPKGRSSIGDMQWDIFVQGNIVNDNDYIEKNNHKSINTGARLPITKNI
VOQGVSVINDKSYEGLSKNSILSGLSNSESFLWGNAKNYOSISYTDGFSLSF
YHNDKVDNCGRNYNAGWCYSYSASUSIPLGLWITSLIGSDTISVESYKSHLSE
YGFYNQNIYKGRQWLTSTSLKWDYNFPAIGIYNSQRLDKGQYISVITIR
ASENSFLDINGSYNGRYNSNELFVGYMTSTNRGDYHEAGMRNKNRHNABGLS
GRINRFGLDNGFSMKNRNTNSTHSLTGGYNSFALTSDDGYWGGSTAGLTLAG
GIIVKNSNDTKNIVVTGFLGYDSI GSNNDNAPI PVPAITPASLIIEDNNYGDNNIS
ILAPTNNMPLPGNVYVPEIETKVSVYIGRGFDNCPPLSGAHVLPNPHVLLDRDG
GRSFYTGNEKTLFLKGRTIYTCQLGKKNVHKGI VFGVDVICDINSSTSLPDEFVN
PRVQDLLAKNDKG"
2170. .2235
/gene="cfa c"
mat_peptide
2236. .4776
/gene="cfa c"
/products="colonisation factor antigen c"
4783. .5865
/gene="cfa e"
CDS
4783. .5865
/gene="cfa e"
/Note="unpublished sequence"
/codon_start=1
/transl_table=11
/product="colonisation factor antigen e"
/protein_id="AAC41417.1"
/db_xref="GI:145511"
/translations="MKKILIFILFRSSGEFTFAVSADKNPGENMNTNIGPHDRGGS
SPIYNILNSYKREYSHHLYDRKSFCLSSQNTLNGACPSDAPGPAIDGENTIL
QPTKRSILKRELIQYQKQFLKNCPSKLALNSHFQCNREQSGATLSLYIPAG
ELNKLFGGVNVAKLNVKRRDYDTYGTITINILNIDKGNIIQWLPQFKNARVD
LNLRPTGGGYIIGNSVDMGCFYDGYSTNSSLIRIFODDNSKSDGKPYLKKINDSKE
LVYTLISLLAGKNIPTNGQALNITASLETNKNRITAVTMEPISVPVLCWPGRLQLD
AKVNPEAGQYMGNIKITFTPPSQTL"

```

```

sig_peptide 4783. .4842
/gene="cfa e"
mat_peptide 4843. .5862
/gene="cfa e"
/products="colonisation factor antigen e"
complement(6388. .6822)
/gene="cfa d"
complement(6388. .6822)
/gene="cfa d"
/note="truncated CfaD-like protein with a stop codon and
frameshift deletions between bp 6029 and 6824."
/codon_start=1
/transl_table=11
/product="colonisation factor antigen d"
/protein_id="AAC41418.1"
/db_xref="GI:145512"
/translations="MDPKYTERKEMIKINNMIHIYVLTNSCIMDIYESEKIPCS
SNRLVFLBEGVNISVRIQKILSEKPYIAFLRNGDILRLKNALMIYGMSKIDINDC
RNMRSRKIIKTEVDKLLDLVKININSYDDSVFISLLIYLISKI"
ORIGIN
Query Match 78.5%; Score 852.4; DB 1; Length 7174;
Best Local Similarity 87.2%; Pred. No. 4.7e-173;
Matches 947; Conservative 0; Mismatches 136; Indels 3; Gaps 1;
QY 1 ATCAATAAGATTTTATTTTATTTTACATTGTTTCTCTTCAGTACTTTTACATTGCT 60
DB 4783 ATGAATAAAATTTTATTTTATTTTACATTGTTTCTCTTCAGTACTTTTACATTGCT 4842
QY 61 GTATCGGCAGATAAAATTCGCGAGATGAAAGCATAACTAATAATTTTGGCCGCGCTGAC 120
DB 4843 GTATCGCGAGATAAAATTCGCGAAGTGAAACATGACTAATCTATTGTCCTCCCATGAC 4902
QY 121 AGGAACGAATCTTCCCCCACAATAATATTAATAAACAACATATTACAGATACAGTAA 180
DB 4903 AGCGCGGATCTTCCCCCATATATAATCTTAAATCTTATCTTACAGCATCAATGGA 4962
QY 181 AGTCATACCTCTGTATGATAGTAGTACTTTTATGTTGTTCTCTTCACATACACTTAAT 240
DB 4963 AGCCATCATCTGTATGATAGTAGTACTTTTATGTTGTTCTCTTCAAAATACACTGAAT 5022
QY 241 GGAGCATGCCCAACCACTGAGAACTCTAGCAGTTCATCGGTGAGCGGTGAAAACAATATA 300
DB 5023 GGAGCATGCCAAGCAGTAGTGCCTCGCCTGGCACTGCTACAATTGATGGCGAAAACAATATA 5082
QY 301 ACATTACAAATTTACGAAAAAAGAAATTAATAAATAAAGAGAGCTACAAAATAAAGGCTAT 360
DB 5083 ACATTACAAATTTACGAAAAAAGAACTCTAATAATAAAGAGAACTGCAAAATTAAGGCTAT 5142
QY 361 AACCAATTTTGTTCAAAAGTGTAACTCCCACTCCGCGCTAACACTTAACCTAAGCTCAT 420
DB 5143 AACCAATTTTGTTCAAAAGTGTAAATTTGCCCATCTAACTAGCAGTAACTCATCTCAT 5202
QY 421 TTTAACTGTAATAAAAAACGGCGCTTCAGGTGCAAGTTTATTTATATATATTCCTGCTGGC 480
DB 5203 TTTCAATGTAATAGAGACACAGCTTCAGGTGCTACTTTATTCGTTATACATACCAGCTGT 5262
QY 481 GAACATAAAATTTGCTTTTGTGTGATCTGCGGATGCTACTCTGAAGTTAAGAGTAAAA 540
DB 5263 GAATTAATAAATTTACCTTTTGGGGGGGTCTGGAATGCCGTTCTGAAGCTAAATGTAATA 5322
QY 541 AGACGATAGTCAGACCTATGGACCTTACACATATAAATATACATTTAAATTAACCTGAT 600
DB 5323 AGACGATAGTACAAACCTATGGGACTTTACACATATAAATACATCACATTAAATTAACCTGAT 5382
QY 601 AAGGGAATATTTCAGATATGTTTACCTCAGTTCAAAAGTCACCTCCGCTCGATCTTAAC 660
DB 5383 AAGGGAATATTTCAGATATGTTTACACACTTCAAAAGTAAACCTCGTCTCGATCTTAAC 5442
QY 661 TTGCGTCCAACTGGTGGGGGCACATATATTGGAAGAAATTTCTGTGATATCTGCTCTTAT 720
DB 5443 TTGCGTCCAACTGGTGGTGGTACATATATCGGAAGAAATCTGTGTGATATCTGCTCTTAT 5502

```

```
QY 721 GATGGATATAGTACTAACAAGCAGCTCTTTGGAGATAGATTTTCAGGATACAAATCTCTAAA 780
Db 5503 GATGGATATAGTACTAACAAGCAGCTCTTTGGAGATAGATTTTCAGGATACAAATCTCTAAA 5562
QY 781 TCTGATGGGAATTTTATCTAAGGAATAAATGATGACACCAAGAAATTCATATCT 840
Db 5563 TCTGATGGGAATTTTATCTAAGGAATAAATGATGATCCAAAGAACTTGATATACAT 5622
QY 841 TTGTCTACTCTCTTGGCGGGTAAAAAGTTTAACTCCCAAAATGGAACGTCATTAAATAT 900
Db 5623 TTGTCTACTCTCTGGCAGGTAAAAATTTAACHCAACAAATGACAGGCAATTAATAT- 5681
QY 901 GCTGACGAGCTTCTCTGGAACAAACATGGAATAGAAATTCACACTGTACCATGCCAGAA 960
Db 5682 --TAACTACTCTCTCTCGGAACAAACATGGAATAGAAATTCACACTGTACCATGCCAGAA 5739
QY 961 ATCAGTCTCCCGTGTGTGCTGCTGCGTTCGACGTTTCGAATGATGCAAAAGTGGAAAT 1020
Db 5740 ATCAGTCTCCCGTGTGTGCTGCTGCGTTCGACGTTTCGAATGATGCAAAAGTGGAAAT 5799
QY 1021 CCGAGGCTGCAATATATGGGTAAATTAATGTTACTTTCACACCAAGTACTCAAAACA 1080
Db 5800 CCGAGGCTGCAATATATGGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5859
QY 1081 CTCTAG 1086
Db 5860 CTCTAG 5865

RESULT 6
LOCUS EOCOCCD
DEFINITION E.coli cooC and cooD genes.
ACCESSION X76908
VERSION X76908.1 GI:488735
KEYWORDS cooC gene; cooD gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 5336)
AUTHORS Froehlich,B.C., Karakashian,A., Melsen,J.R., Wakefield,J.C. and Scott,J.J.
TITLE CooC and CooD are required for assembly of CSI pili
JOURNAL Mol. Microbiol. 12 (3), 387-401 (1994)
MEDLINE 94344028
PUBMED 7915003
REFERENCE 2 (bases 1 to 5336)
AUTHORS Scott,J.J.
TITLE Direct Submission
JOURNAL Submitted (24-DEC-1993) J.J. Scott, Emory Uni. School of Medicine,
Dep. of Microbiol. and Immunol., 1510 Clifton Rd, Atlanta 30322,
Georgia, USA
COMMENT Entry cited as #76908 in the journal.
FEATURES
source
location/Qualifiers
1..5336
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="LMC10"
/db_xref="taxon:562"
/clone="pEU494"
76..79
/citation=[1]
95..2703
/gene="cooC"
95..2703
/gene="cooD"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CooC"
/protein_id="CAA54225.1"

RBS
28.6%
Query Match
Best Local Similarity 58.8%
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;
Pred. No. 1.1e-56;

Gene
76..79
/citation=[1]
95..2703
/gene="cooC"
95..2703
/gene="cooD"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CooC"
/protein_id="CAA54225.1"

CDS
76..79
/citation=[1]
95..2703
/gene="cooC"
95..2703
/gene="cooD"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CooC"
/protein_id="CAA54225.1"
```

```
/db_xref="GI:488736"
/db_xref="SPTREMBL:Q47114"
/translation="MIGGSKVWIVLVILGSSGSEFASKYNLVIPESFRDLWGEQD
ELLVRLVQCQSLGVRHRIKSTPTTVAPEPNDLLDKIEINKEADLRVLMRGSFORNG
NMSCOGYTGONNVCNKINTVAIVDDVENVLNLFIENRFLASGENSDSYQPSKNTK
KAFIHSOTINLSDTGNVENLSIVGSGLITNSVAILGWAANYRKYKSYNKEOSIN
SLYFRHDEKFNFYQIGRIDRDLSSQSGGNEFDLLPVFDIYGRAGTTQSYIKNTG
KSAFVETIMLTHFSRVEAYRQQLIGVWLDAGISELTERPDGNYDLKLKIFEQE
QIVRBEIVFPNKGSSGIDTWDVFWQADIIINDGRVYVEKQKNHKSAINSLRPLT
RNLVQLGAGVINDKNYVETGLMNSGLDLSNKSFTFLFGDTHGNTQNTSYTDGF
SISVYHNDKRVDDCGKDNMGWGCYVESASLSIPVKGNSTLAYSNYTSVRYD
AVSVYVYVYKRTKRMOLTASTVVRMGDYNIMPTIGVNSBEQKADKGGVLSLTL
TRVDGKSLNAGSYNYSRCNTSNDAFVEGHLVSDTNVYSRELSARVSGSVYTBGG
TVSGIRNRFGDNGTLINNNKRSHTTHSLTAGYSSFALTDDGLYWGSGASGLTNL
SGGIVRVKSNEDSELNLYKSGSYGHSYNSDSIFIPVPAKQASLTLENTNKSNN
IDVLAFTNTFFMLPSGVYPIVDSANVSFTYVGRGVKGRPLSGAYIIINAQNIILDE
NGGFFPESSENEKELFLKDKTIYSCSLDRSEMRNGIAFVGEVACNSTIKELLPEKL
TNSRTHDLLAYNQDTE"
83..156
/gene="cooC"
/citation=[1]
157..2700
/gene="cooC"
/product="CooC"
/citation=[1]
2889..2892
/gene="cooC"
/citation=[1]
2700..3791
/gene="cooD"
2700..3791
/gene="cooD"
/citation=[1]
2700..3791
/gene="cooD"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CooD"
/protein_id="CAA54230.1"
/db_xref="GI:488737"
/db_xref="GOA:Q47115"
/db_xref="SPTREMBL:Q47115"
/db_xref="SPTREMBL:Q47115"
/translations="MKKIFILSIIFSAVVSAGRYPETTVGNLTKSFQAPRLDRSVQS
PIYNTNVAAGYSLSHSYDRIVFLCTSSNPVNGACPTIGTSVGYQVGTTLTLOFT
EKRLIKRNINLGNKKPIWENQSCDFSNLWLNKSWSCGAGNANGTLINLYIPAG
EINKLPFGGIWEATLILRLSRYGEVSTHYGNITVITDLTDAGNIQWLQPHFNP
RVDNLNRPIGNYKSGNSLDMCFYDGYTSDSMVIRKPDDEPTNSSEVNIYKIGGT
EKLFPVSLIMGEKIFYPVNGQSTFINDSVLETNWNRVTAVAMPEVNVPLCWPAEL
LLNADVNPADAGQYSQIYITFTPSVENL"
2700..2753
/gene="cooD"
/citation=[1]
2754..3788
/gene="cooD"
/product="CooD"
/citation=[1]
476..5336
/notes="bases 1-887 of IS2"
/insertion_seq="IS2"

repeat_region
sig_peptide
mat_peptide

ORIGIN
Query Match
Score 310.8; D3 1; Length 5336;
Best Local Similarity 58.8%; Pred. No. 1.1e-56;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;

QY 7 AAGATTTTATTTTATTTTACATTTGTTTCTCTCTCAGTACTTTTTTACATTTGCTGTATCG 66
Db 2703 AAAAGATATTTATTTTCTATCATATTTTCTGCGTGTGCTAGTCCGCGGATAC 2762
QY 67 GCAGATAAATTCGCGAGATGAAGCAATACTAATAATTTTTGGCCCGCGTGCAGAAC 126
Db 2763 CCGGAATACTACAGTAGGTAATCTGACGAGAGTTTTCAGCCCGCTCTCTGGATAGAGC 2822
QY 127 GAATCTTCCCCCAACATAATAATATAATAAACATATTACAGCATACAGTGAAGTCAAT 186
```

Db 2823 GTACAATCCCAATATAAATCTTTACGATCATGCTGGCTGGATATAGTTTGGATCAT 2882
Qy 187 ACTCTGTATCAGGATGACTTTTATGTTTGTCTTCTCAAAATACACTTAATGGAGCA 246
Db 2883 AGCTTATATGACAGATTTGTTTATGATCATCTCTGCGAATCCGGTAAATGGTGT 2942
Qy 247 TGTCCACAGTGAGAAATCCTAGCAGTTCATCGGTGAGCGGTGAACAATAATATACATTA 306
Db 2943 TGGCCAAACCAATGGAACATCTGAGTTCAAT-----ACGGTACTACAACCAATAACCTTG 2996
Qy 307 CAATTTACGGAAGAAAGAGAGTTTAAATAAAGAGAGCTACAAATTAAGAGCTATAACAA 366
Db 2997 CAGTTTACGAGAAAGAGAGTCTGATAAAGAGAAATTAATTAATCTGCAGGTAAATAGAA 3056
Qy 367 TTAATCTTCAAAAGTCTTAAGTCCCATCCGCTAACACTT-----AACTCAGCTCAT 420
Db 3057 CCNATATGGAGAAATCAGAGTTGCGACTTTAGCAATCTAATGGTGTGAATCGAAGTCT 3116
Qy 421 TTTAACTGTAAATAAAACG---CGGCTTCAGGGGCAAGTTTATATATATATATCTCTGCT 477
Db 3117 TGGAGCTGTGGGCTCACGGAATGCTAAGCGAACAATCTTAATCTGATATCCCTGCA 3176
Qy 478 GCGCACTAAATATGCTTTGCTTGGTCTATCTCGGATGCTACTCTGAAGTTAAGAGTA 337
Db 3177 GGAGAAATCAACAAATGCTTTTGGAGGATATGGAGGCAACTCTGATCTTACGCTTA 3236
Qy 538 AAAAG-----ACGATATAGTGAGACCTATGGAACCTTACACTATAATATCACTATT 588
Db 3237 TCAGATATGCGGAAGTCACTASCACCAATAGCGCAATTAATACCGTAAATATACGGTT 3296
Qy 589 AAATTAACATGAAGGAANAATTCAGATATGTTTACCTCAGTTCAAAAGTGAAGTTCGCTGC 648
Db 3297 GATTTAACTGATAAGGAATATATTCAGGTATGGCTTCCAGGGTTTACAGCAACCCCGCT 3356
Qy 649 GTCGATCTTAACCTGCTTCAACTGCTGGGGGACATATATGGAAGAAATTCGTGTAT 708
Db 3357 GTAGACCTGAATCTGGCCCTATCGGTAATATAAATATAGTGGTATTAATCACTCGAC 3416
Qy 709 ATGTGCTTTTATGATGATATAGTACTTAACAGCAGCTCTTTGGAGATAAGATTTCCAGAT 768
Db 3417 ATGTGTTTCTATGATGATATAGTACAAACAGTATAGCATGATGAATTAAGATTTCCAGAT 3476
Qy 769 AACAACTCTAAATCTGATGCGCAATTTTATCTAAGGAAAAATAATGATGACACCAAGAA 828
Db 3477 GATAATCTTACCAATCATCTGAATATAAATCTTATAAGATA---GGGGCCTGAAAAA 3533
Qy 829 ATTGCATATACTTTGTCACCTCTCTGGCGGTAAGATTTAACTCCAACAAATGGAACG 888
Db 3534 TTACCATATCTGTTTCACTGCTTATGGGAGAAAAATAATTTATCCAGTGAATGGTCAA 3593
Qy 889 TCATTAATATTTGCTACGCGAGCTTCTCTGGAACAAACTGGAATAGAAATTAACAGTGTG 948
Db 3594 TCATTTACATCAATGACATGATGATCTGAAACAACTGGAATCGAGTAACCGAGTT 3653
Qy 949 ACCATGCCAGAAATCAGTGTTCGGTCTTGTGTGGCTCGGAGCTTTGCAATGGATGCA 1008
Db 3654 GCTATGCCGGAAGTAAATGTTCCAGTATTAATGCTGGCCAGCAAGATGCTAATAAATGCT 3713
Qy 1009 AAGTGGAAATCCGAGGCTGACAAATATATGGTAAATTAATGTTACTTTACACCA 1068
Db 3714 GATGTAATGCTCCGATGAGGACAGTATTCAGGACAGATATATAAATTAACATTTACACC 3773
Qy 1069 AGTAGTCAAA 1078
Db 3774 AGTGTCAAA 3783

RESULT 7
AX741413
LOCUS
DEFINITION Sequence 3 from Patent WO03022306.
ACCESSION AX741413

AX741413.1 GI:30524203
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1
Turner,A.K., Greenwood,J., Stephens,J.C., Beavis,J.C. and
Darsley,M.J.
Bacterial vaccine
Patent: WO 03022306-A 3 20-MAR-2003;
Acambis Research Limited (GB)
Location/Qualifiers
1..5336
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
85..2703
/note="unnamed protein product; cooC"
/codon_start=1
/transl_table=11
/protein_id="CAD90906.1"
/db_xref="GI:30524204"
/db_xref="RMTREMBL:CAD90906"
/translation="MIGSSSGFASKYNLVLSVLIGSSSGFASKYNLVLPESPRDLWGQD
ELLEVLYGQSIGVHRKSTPTTAVFSPDNLLDKIEINKKKEADRLVIRGSRFQNG
NMSQGTGQNNCKYIKINTVAVDDVENLVNLFIGNFLASGENDSDYYQESKNTK
KAFTHQNTLNQNTGNYENLISVGTSLGTDNSVAILGMAANYRKYNYNEQSN
SLYFHPDKNFYQGLRIDRSDLSQSGCNFNFDLLPVDIYGMAGTQSYKNTG
KSAAPVTMLTHFSRVAAYRNGQLLGVWLDAGISELDERLPDGNVCLKIKIPEO
QVAREIYVFNKSGSSIGTDHNDVFQAGDIINDAGRYVEKQKNHKSANSGRLDIT
RLAVOLGGAVIDNNKYEIETGLMNSGLDGLNSKFTFLFGDDTHGNTQNVSTGOF
SLSPFYHDKVDCCQDYNMGWGSYESASLSIPVKGNMSTLAFNTQNTSYRYD
AYSEVPPYKYKTRKQWLTASTVVRWGDYIMPTIGVNSGKQWAKGGVSLTL
TADGGKSNAGSYNSVSGNTSNDAPVEGHLVSDTVNSYRELSPVSGRNVYTEGG
VSGRINRRPGLDGLNVAKNKSHDTTSLTAGYSSSALUTDGLYWGSGASGLNL
SGIVRVKSNEDBSLLNVKGSYGHSLGSDSLFIPVPAALMQASLTTEENTKSKQ
IDVLAIRPTFFMLPGSVYPIDVANVSPTVGVGVDPGRPLSGAYILNAQNVILDE
NGSFEFSESENEKELFLKDKDTYSCSLDRSEMRNGIAFVGEVACNSTIKELLPEKLV
TNSRIHDLIAVYNDQTE"
2700..3791
/note="unnamed protein product; cooD"
/codon_start=1
/transl_table=11
/protein_id="CAD90907.1"
/db_xref="GI:30524205"
/db_xref="RMTREMBL:CAD90907"
/translation="MKKIFILSIIPSAVVSAGRYPETTVGNLTJGSPQAPRLDSVQS
PYNIPTNVAGYSLSHSLYDRIVECTSSSNP'VNGACPT'GTS'G'QYCTTTITLOFT
EKRLIRKINILAGNKKPIWENQSCDFSNLMVNSKSWSCGARGNANGILLNLIYIAG
EINKLPFGGIWEATLILRLSYRGEVSTHYGNVNIIVDLTDKGNIQWLPGFHNE
RYDLNLRPTGNYKYSNSLDMCFYDGYSTNSDSMVYKFDNDKPTNSSEYNLYKIGST
EKLPAVSLIMGEKIFYPVNGQSFINDSVLETNWNRVTAVAPVNPVPLCWPAEL
LLNADVNPADAGYSGQIYITE'PSVENL"
Query Match 28.6%; Score 310.8; DB 6; Length 5336;
Best Local Similarity 58.8%; Pred. No. 1.1e-56;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;
QY 7 AAGATTTTATTTATTTTACATTTGTTTCTTCTCAGTACTTTTACATTTGCTGATCG 66
Db 2703 AAAAAGATATTTATTTTGTCTCTATCATATTTCTCGCGTGGCTAGTGGCGCGATAC 2762
QY 67 GCAGATAAATTCGCGAGATGAAGCAATAATAATTTTGGCCCGCGGTGACAGAAC 126
Db 2763 CCGGAACACTACAGTAGTATCTGACGAGAGTTTTCAGACCCCTCGTCTGGATAGAC 2822
QY 127 GAATCTTCCCCCAACATATATATTAATAACATATTTACAGCATACAGTGAAGTCAT 186
Db 2823 GTACAATCAACATATATAACATCTTACGAATCATGTGCTGGATATAGTTGAGTCAT 2882
QY 187 ACTCTGATGATAGGATGACTTTTATGTTTCTTCTCTCACAAATACACTTAATGAGCA 246

Db 2883 AGCTTATATGACAGGATGTTGTTTTTATGTATACATCTCGTGAATCCGGTAAATGCTGCT 2942
Qy 247 TGTCACACAGTGAATCCTAGCAGTTCATCGTCAGCGGTGAACAAATATAAACA 306
Db 2943 TGCCCAACCATTTGAACATCTGGAGTCAAT-----ACGTTACTACACCAATACCTTG 2996
Qy 307 CAATTTCAGGAAAAAGAGTTTAATAAAGAGAGCTACAAATTAAGCGCTATAACAA 366
Db 2997 CAGTTTACAGAAAAAGAGTCTGATAAAAGAAATATTAACTCTGACGCTAATAAGAA 3056
Qy 367 TTATGTTCAAAGTGTAACTGCCCCCTCCGCTACACTT-----AACTCAGCTCAT 420
Db 3057 CCAATATGGGAGATCAGAGTGGCACTTTAGCAATCTAATGTTGTTGAATTCGAAGTCT 3116
Qy 421 TTAACTGTATAAAGACG--CGCTTCAGGTGCAAGTTTATATTTATATATATTCCTGCT 477
Db 3117 TGGAGCTGTGGGCTCAGGAATGCTAAAGGAACATTTTAATCTGTAATCCCTGCA 3176
Qy 478 GGCGAACTAAAAAATTTGCCCTTTGGTGGTATCTGGGATGCTACTCTGAAGTTAAGAGTA 537
Db 3177 GGAGAATCAACAAATTTGCCCTTTGGAGGATATGGGAGCAACTCTGTGATTTACGCTTA 3236
Qy 538 AAAG-----ACGATATAGTGAGACCTATGGAACCTTACACTATAAATATCACTATT 588
Db 3237 TCAAGATATGCGAAGTCAGTAGCACCCATTTACGGCAATTTATACCGTAAATATTACGGTT 3296
Qy 589 AAATTAAGTATAGGGAATATTTCAGATATGTTTACCTCAGTTCAAAAGTGAGCTCGC 648
Db 3297 GATTTAACTGATAAAGTAATTTTCAAGTATGCTTCAAGGTTTCACAGCAACCCGCT 3356
Qy 649 GTGATCTTAACTTSCCTCCAACCTGGTGGGGGCACATATATTTGGAAGAAATTCGTGAT 708
Db 3357 GTAGACCTGAATCTGCCCCCTATCGGTAATATAATATAGTGTAGTAAATCACTCGAC 3416
Qy 709 ATGTGCTTTTATGATGATATAGTACTTAACAGCAGCTCTTTGGAGATAAGATTTTCAGAT 768
Db 3417 ATGTGTTTCTATGATGATATAGTACAAACAGTATAGCATGTAATAAAGTTCCAGAT 3476
Qy 769 AACATCTAAATCTGATGGGAATTTTATCTAAGGAAATATAATGATGACACCAAGAA 828
Db 3477 GATTAATCTCAACATTTATCTGATATATAATCTTTATAGATA---GGGGGCACTGAAAA 3533
Qy 829 ATTGCATATATCTTGTCACTTCTCTTGGCGGTAAAGTTTAACTCCACAAATGGAACG 888
Db 3534 TTACCATATGCTGTTTCACTGCTTATGGGAGAAAAATATTTTATCCAGTGAATGGTCAA 3593
Qy 889 TCATTAATATGCTGACGAGCTTCTCTGGAACAACTGGAATAGATTAAGTATACAGCTGTC 948
Db 3594 TCATTTACTATCAATGACAGTGTACTCGAAACAAACTGGAATCGAGTTAACCGCAGTT 3653
Qy 949 ACCATGCGAGAAATCAGTGTTCGGGTGTTGTTGGCTGACGCTTTCGAATTTGATGCA 1008
Db 3654 GCTATGCGGAGTTAATGTTCCAGTATATGCTTGGCCAGCAAGATTCGTTAAATGCT 3713
Qy 1009 AAAGTGAAGAAATCCGAGGCTGGACAATATATGGGTATATTAATGTTTACATTCACACCA 1068
Db 3714 GATGTAATGCTCCGATGCGAGGACAGTATTCAGGACAGATATATATAACATTTACACCC 3773
Qy 1069 AGTACTCAAA 1078
Db 3774 AGTGTGGA 3783

RESULT 8
LOCUS AY216491
DEFINITION Escherichia coli strain DS37-4 plasmid CS17 gene locus, complete
ACCESSION AY216491
VERSION AY216491.1 GI:28932774
KEYWORDS Escherichia coli
SOURCE

ORGANISM Escherichia coli
REFERENCE 1 (bases 1 to 6956)
AUTHORS Biediger, W. and Wolf, M.K.
TITLE CS17 of Enterotoxigenic Escherichia coli
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6956)
AUTHORS Biediger, W. and Wolf, M.K.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2003) Enteric Infections, Walter Reed Army
Institute of Research, 503 Robert Grant Ave, Silver Spring, MD
20910-7500, USA

FEATURES
source
1..6956
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="DS37-4"
/db_xref="taxon:562"
/plasmid="unnamed"
/notes="naturally-occurring plasmid"
166..879
/gene="csbB"
166..879
/gene="csbB"
/notes="CS17 fimbrial accessory protein"
/codon_start=1
/transl_table=11
/product="CsbB"
/protein_id="AA060095.1"
/db_xref="GI:28932775"
/translations="MRKLSELLMIFPVYKAMFVYPISEIKGASEIIRIYKSKD
TQYIYAKVGVGPKDEYEDAPNWEAGLVVTPARVILFSGNSEVRLTQINAPT
REYVYFESVPEACENSGNKTLDLSVNVYIYALRILPINKNASVISPSN
NIIKNTGNVRLGIMNVFFCKSSINDDCVKSNKNVPGYLLDTMIYNDGYSHUL
DTWDESSNKTDLNLIQVPI"
933..1439
/gene="csbA"
933..1439
/gene="csbA"
/notes="CS17 fimbrial subunit; structural subunit of CS17
pilus"
/codon_start=1
/transl_table=11
/product="CsbA"
/protein_id="AA060096.1"
/db_xref="GI:28932776"
/translations="MKLKITIGAMAVATLPAATMAAVERNIITVRSVDPKLLDQAD
GTSLPDSIALTYSSANNREVSILNTAHTNDKSKGVVVKLSAPVLINMKNSQIP
MKVTLGKTLNTDTBFTVTLNFGTSGVENVSTQQLTTHADTQGTAPAGNVQGI
SLIMTQKT"
1506..4124
/gene="csbC"
1506..4124
/gene="csbC"
/notes="putative fimbrial usher"
/codon_start=1
/transl_table=11
/product="CsbC"
/protein_id="AA060097.1"
/db_xref="GI:28932777"
/translations="MYSKPFSLIVLPFSSPLGPFADKNKITNIPESFRLWGEOD
GFLEVPYGGSLGVHRIKLTPTVAPESPDLIDKIENKEKADLPLVMKGSFORNG
NMSCGYTGQNNCNYIKNTIVAVDDVENVLNFIENFLASGENDSDVYQPSKNK
KAFTHQITNLSDTGNYENLSIVTSGSLGIDNSYALGWAANNRKYKSNYNHQSIN
SLVFRHDFEKNFYQLGRIDTDLSSQSGGNFDFLLPVPDIYGMRAGTQSYIKNTG
KSVASPTVIMLTHFSRVEAVRNGQLLCVWYLDAGISLDTERLPDGNYDLKLIKEQ
QIVREELVPNKSSSISGDTHWDFVQAGDIINDNGRYVEKQKHKAINGSLRLPT
RNLAUQLGAVINDKNYETGILWNSFLDGLSINSKFTLFGDDTHGNYQNVSTDFG
NLSPFHNDRXVINDCGNDYNMGWGCYSISASISIPVKGWRSILAYNTYSTSVRYD
AVSEYFYSYIKGTKEWQLTASTVVRWGDINIMPTIGVYNSEKQWADKGGYLSLIL
TRVDGKSLNAGSYNYSRGNVTSNDAFVEGRIVSDTNVSYRELSAHSVGNRYTEGG
VSGRIINRFGDLNGLTSLVNKRKSHDTHSLTAGYSSSFALTITDGIYWGGSASGLTNL

SGGIVRVKSNDEBELLVNKGSSYXNYSLGNSDLFIPVAPALMOASLTIKENTNKSXN
 IDVLAPKNTFMPGPNVPIDVANVSFTVVGKGVKRPPLSGAVILNQNVLDE
 NGSFSSSENKELFLKDKRTIYSCSLRSEMRNGIVFVGEVACNSFKELLPKLV
 TNSRIHDLAVNQDTE"
 4121..5212
 /gene="csbp"
 /cds="csbp"
 4121..5212
 /gene="csbp"
 /note="CS17 assembly protein"
 /codon_start=1
 /transl_table=11
 /product="csbp"
 /protein_id="AAO60098.1"
 /db_xref="GI:28932778"
 /translation="MKKIFILSIFSVAGRYPETVGNLTKSFQAPQDRSVQS
 PIYNIETHVAGYSLSHLYRIVFLCTSSPNVNGACPNRGTSQVGTITTLQFT
 EKRSLIKENILAGNKKRIWENQSDTSLNVLNKSWSGANGKANTLNLXIYAG
 FINKLPFGIWEATILILSLRYGVDSYHNYTVNIIVDLTDKNGICVWLPGFHNP
 RVDLNLHIGNKYSGSLSLDMCFYDGYSTNSDNLVIFKODNPYTSSEVLYKIGET
 EKLPAVSLLMGEKILFHPVNGQSFTINDSSVLEINWNRVIAVMPVNPVLCWPARL
 LLNADVNAPDAGQVSGQIYIUTFPSVNL"

ORIGIN

Query Match 28.5%; Score 309.6; DB 2; Length 6956;
 Best Local Similarity 58.2%; P: 0; Mismatches 429; Indels 27; Gaps 4;
 Matches 634; Conservative 2

4124 AAAAGATATTTATTTTGTCTATCATATTTCTGGGTGTCAGTCCGGGGGATAC 4183
 7 AAGATTTTATTTATTTTACATGTTTTCTCTTCAGTACTTTTTACATTTGCTCTATCG 66
 67 GCAGATAAAATCCCGGAGATGAAGCATACTAATATTTTTCGCCCGCGTGACAGGAAC 126
 4184 CCGGAATCTACAGTAGTGAATCTGACGAGAGTTTCAAGCCCTCGTCAGGATAGAAGC 4243
 127 GAATCTTCCCCAACAATATATTAATAACCATATTTACAGCATACAGTGAAGTCTAT 186
 4244 GTAACATCCGAAATATAACATCTTTACGATCATGTGGCTGGATATAGTTGAGTAT 4303
 187 ACTCTGTATGATAGGATGACATTTTATGTTTGTCTTCAAAATACACTTAATGAGCA 246
 4304 AGCTTATATGACAGATTTGTTTTTATGATACATCTCTGTGAAATCCGGTTAATGTTGT 4363
 247 TGTCAACAGTGAATCTTACGAGTTCATCGGTACGGGTGAACAAATATACATTA 306
 4364 TCCCCAAACCGTGGAAACATCTGAGCTCAAT-----ACGGTACTACACCAATACCTTG 4417
 307 CAATTTACGGAATAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTATATAACAA 366
 4418 CAGTTTACAGAAAAAAGAAAGTCTGATAAAGAAATATTAATCTTGCAGGTATTAAGAA 4477
 367 TTAATGTTTCAAAAGTGTAACTGCCCATCCGGCTTAACACTTAAGTCAAG-----CT 417
 4478 CCAATATGGGAGAAATCAGAGTTTCGACACTAGCAATCTAATGGTGTGAATTCGAAGTCT 4537
 418 CATTTTAAGTAAATAAACCAGCTTCAGGTGCAAGTTTATTTATTTATTTATTTCTGCT 477
 4538 TGGTCTGTGGGGCTAAACGGAATGCTAACGGAACACTTCTAATCTGTATATCCCTGCA 4597
 478 GCGGAACATAAAATTTGCTTTTGGTGGTACTCGGATGCTACCTGAAGTAAAGATTA 537
 4598 GGAGAAATCAACAATATGCTTTTGGAGGATATGGAGGCACTCTGATCTTACGCTTA 4657
 538 AAAAG-----ACGATATAGTCAGACCTATGGAACCTTACACTATAAATATCACTATT 588
 4658 TCAAGATATGGCAAGTCAGTACGACCCATTACGGCAATATATACCGTAAATATACGGTT 4717
 589 AAATTAAGTATAAGGGAATTTTCAATATGTTTACCTCAGTTCAAAAGTGAAGCTTCCG 648
 4718 GATTTTAAGTATAAAGGTAATATTCAGGTATGGCTTCCAGGGTTTACAGCAACCCGGCT 4777
 649 GTCGATCTTAAGTGGTCCCACTGTGGGGGCACATATATTTGGAAGAAATTTCTGTGAT 708

4778 GTAGACCTGAATCTGCACCCCTATCGGTAATATAAATATAGTGGTAGTAAATTCACCTCGAC 4837
 709 ATGTGCTTTTATGATGATAGTACTACTAAGCAGCTCTTTGGAGATAGATATTTCCAGGAT 768
 4838 ATGTGTTTCTATGATGATATAGTACAAACAGTGTATGATGATGATGATGATGATGATGAT 4897
 769 AACATCTCTAAATCTGATGGGAAATTTTATCTTAAGGAAATATAAATGATGACACCAAGAA 828
 4896 GATTAATCTCTACTATCTATCTGATATATATTTTATAAGATA--GGGGGCACTGAAAA 4954
 829 ATTGCATATCTTTTCTACTTTCTCTGCGGGGTAAAAAGTTTAACTCCCAACAAATGACG 888
 4955 TTACCATATGCTGTTTCTACTCTTATGGGAGAAAAATATTTTATCCAGTGAATGCTCAA 5014
 889 TCATTAAATTTGTCGACGAGCTTCTCTGAAACAAACTGGAATAGAAATACAGCTGTC 948
 5015 TCATTACTATCAATGACAGTAGTGTACTCGAAACAAACTGGAATCGAGTAAACCCAGT 5074
 949 ACCATGCCAGAAATCAGTGTTCGGGTGTTGTTGCTGCTGACCTTGCATATGGATGCA 1008
 5075 GCTATGCCGGAAGTTAATGTTCCAGTATATGCTGCGCCAGCAGATGCTGATTAATGCT 5134
 1009 AAAGTGGAAATCCCGAGGCTGCAATATATATGTTGTTTAACTTCTTACTTTCACACA 1068
 5135 GATGTAATGCTCCGATGCGAGCAGTATTCAGGACAGATATATAACATTTACACCC 5194
 1069 AGTAGTCAAA 1078
 5195 AGTGTGCAAA 5204

RESULT 9
 ECOTABCD
 LOCUS E.coli cotA, cotB, cotC and cotD genes. linear BCT 13-MAR-1996
 DEFINITION
 ACCESSION Z47800
 VERSION Z47800.1 GI:897725
 KEYWORDS cotA gene; cotB gene; cotC gene; cotD gene.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 5798)
 AUTHORS Froehlich,B.J., Karakashian,A., Sakellaris,H. and Scott,J.R.
 TITLE Genes for CS2 pili of enterotoxigenic Escherichia coli and their
 INTERCHANGEABILITY with those for CSI pili
 JOURNAL Infect. Immun. 63 (12), 4849-4856 (1995)
 MEDLINE 96071908
 PUBMED 7591145
 REFERENCE 2 (bases 1 to 5798)
 AUTHORS Froehlich,B.J.
 TITLE Direct Submission
 JOURNAL Submitted (18-JAN-1995) Froehlich B. J., Emory University,
 Department of Microbiology and Immunology, 1510 Clifton Rd, Atlanta,
 GA, USA, 30322

FEATURES
 source
 1..5798
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="C9if-6"
 /db_xref="taxon:562"
 /clone="pEJ5C06"
 1..369
 /note="99% homologous to last 369 bases of IS3"
 /citation=[1]
 /label=IS3
 369..401
 /note="100% homologous to bases 73-41 of IS1"
 /citation=[1]
 /label=IS1
 422..427
 /label=IS1
 446..451
 /citation=[1]
 -35_signal
 -10_signal

```

RBS      485..488
/citation=[1]
gene      499..1215
/gene="cotB"
CDS      499..1215
/gene="cotB"
/citation=[1]
/codon_start=1
/transl_table=11
/product="CotB"
/protein_id="CAA87760.1"
/db_xref="GI:897726"
/db_xref="SPTREMBL:Q47116"
/translation="MKILLFVLFNFVPAASANFVYPLSKDLOQSGSETIKVFKSK
DYQIKYTKRVINPTEKEQVDIKMDGGHIVPAKVLPAAGASKRIURIEINKKE
QEEVRYVESYKQOQDDIEEKGRVNTDLVNIIYAALIRTSNPQRKIDVSIIES
NNWIKNTGNIRLGIKQDVFLOCTTISINDKCAKFSYRNLYPDMQVDTKLKGGKFSYAV
IDTKDRNENSGELINIKLP"
sig_peptide      499..552
/gene="cotB"
mat_peptide      553..1212
/citation=[1]
/gene="cotB"
/product="CotB"
/citation=[1]
1241..1245
/citation=[1]
1255..1767
/gene="cotA"
1255..1767
/gene="cotA"
/citation=[1]
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CotA"
/protein_id="CAA87761.1"
/db_xref="GI:897727"
/db_xref="SPTREMBL:Q47117"
/translation="MKLNKIIIGALVLSSTFVMSGASAEKNITVTASVDPTIDLQSQD
GTALPSAVNIAYLPQEKRPESARINFTQVHTNNKTKGIQIKLNDNVVMNLDSPKTI
PLEVSPAGTKLSTAITSITADQLNFGAAGVETVSATKELVINAGSQQTQTNIVAGNYQG
LVSIVLTQSP"
sig_peptide      1255..1323
/gene="cotA"
/citation=[1]
/evidence=experimental
1324..1764
/gene="cotA"
/product="CotA"
/function="major pilin subunit for CS2 pili"
/citation=[1]
/evidence=experimental
1822..1824
/citation=[1]
1836..4436
/gene="cotC"
1836..4436
/gene="cotC"
/citation=[1]
/codon_start=1
/transl_table=11
/product="CotC"
/protein_id="CAA87762.1"
/db_xref="GI:897728"
/db_xref="SPTREMBL:Q47118"
/translation="MRANFKITVFIPLIGLGFGTNGLBSKKNIPBEFIDLMEODEL
LEVNLRSLGVRHVLITPTTVFSPVEEILEKINVKQEKEDLRSLILQSYSENGM
SCNGFDEKYSYCNIRDTNVNIVDENNELNIFIGASFLSVQAOQNIIYQKNINSEK
AFIESQTINPESSEGYKSLKGVAGQGLTENSYLVFGMDAIYNSRKYTYKNQSIINN
IYRYDFDKYY"QLGMDRSDLSASSGNFNFMPLPLPIDGFQIGTQSYIKNIEK
SISSPVTMLITRESVYAPRNBELLGWYLNLSGINDLD"SRLPDQSYDLTLKIFEQD
LVREKVPFNKGGASPDQXQDVFPAQAGNIVNNNDSVIERKQTKNKTGINAGITPVR

```

```

sig_peptide      1836..1913
/gene="cotC"
/citation=[1]
1914..4433
/gene="cotC"
/product="CotC"
/citation=[1]
4441..4445
/citation=[1]
4451..5545
/gene="cotD"
4451..5545
/gene="cotD"
4505..5542
/gene="cotD"
/product="CotD"
/citation=[1]
mat_peptide      4505..5542
/gene="cotD"
/product="CotD"
/citation=[1]
ORIGIN
Query Match      27.5%; Score 298.8; DB 1; Length 5798;
Best Local Similarity 58.5%; Pred. No. 4.1e-54;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;
QY      127 GAATCTTCCCAACATAATATATAAATAACCATATACAGATACAGTGAAGTCAT 186
DB      4583 GCATCGTATCCCGCTCATTATATATTCATGACATGTCATCAACAGATCATCTAAAGTGCA 4542
QY      187 ACTCTGTATGATAGGATGACTTTTTTATGTTTCTCTCTCACAAATACACTTAATGGAGCA 246
DB      4643 TCUCITTTTCAGAGGATGACGTTTATGATGTCATCAACAGATCATCTAAAGTGCA 4702
QY      247 TGTCCAAACAGTGAGAAATCCTAGCATTTCATCGGTGAGCGGTGAACAATATATAATTA 306
DB      4703 TGTCCGACAGGAGAAAACCTCCA-----AATCCTCTCAAGGGGAGACTAATATAAGCTA 4756
QY      307 CAATTTACGGAAGAAAGAGAGTTTAATAAAGAGAGCTACAAATTAAGGCTATAAACA 366
DB      4757 ATATTTACTGAAAGAAAGAAAGTCTGCCCAAGAAAACATTAACATTAAGAGATATAAGAGA 4816
QY      367 TTATCTTCAAAGTGTTAACTGC-----CCATCCGGCCTAACACTTAACCTCAGCT 417
DB      4817 TTTTATATGATCAGATAGATGCAATTCATTATGTCGATAAAGATGAATCTCAATTCCTCAT 4876
QY      418 CATTTTAATGTAATAAAACGGGGCTTCAGGTGGAAGTTTATATTTATATATTCCTGCT 477
DB      4877 ACTGTTAAATGTGTAGGTTCAATCAACAGAGAGTAGATTTTCATCTTATATATATCCACAA 4936

```


QY	478	GGCGAACTAAAAAATTTGCGTTTGGTGGTATCTGGGATCGTACTCTGAAGTTAAGAGYA	537
Db	4937	GGTGAATTCATGGGGTCTTAACTGGAGGTATATGGGAGCAACACTAGAGTTACGAGTC	4996
QY	538	AAAAGACGATATAGTGAACCTATGGAACCTTACACTATAATATCACTATTAAATTAAC	597
Db	4997	AAAAGCATTCAGACTATAATCTAAGTGAACCTTACACTATAATATCACTATTAAATTAAC	5056
QY	598	GATAAGGAAATATTCAGATATATGTTACCTCAGTTCAAAAGTGAACGCTCGCTCGATCTT	657
Db	5057	GACAAAGGAAATATTCAGGTCCTGGACACCAAGATTTTATAGCGATCTAGAAATGATCTG	5316
QY	658	AATCTCGCTCAACTCGTGGGGGACATATATTTGGAAGAAATCTGTGTATGTGCTTT	717
Db	5117	AATTTAGCTCCTCAAGGTAATGTTAAATATTTCTGGTAGTAACGCTGTGATGTGCTC	5176
QY	718	TATGATGGATATAGTACTAACAGCAGCTCTTTGGAGATAGATTTCAAGGATAACAATCCT	777
Db	5177	TATGATGGCTATAGTACACTAGTCAAGATATAGAAATGAGGTTTCAGGATGACTCACA	5236
QY	778	AAATCTGATGGGAAATTTTATCTAAGGAAATATAATGATGACACCAAGAAATTTGCATAT	837
Db	5237	ACAGGAAATATAATGAATATAATCTTATAAAACTGGAGAGCCATTAAAAAATTTGCCATAT	5296
QY	838	ACTTTGCTACTCTCTTGGCGGGTAAAGTTTAACTCCACAAATGGAACGTCATTAAT	897
Db	5297	AAATCTTCTCTCTTTTAGGAGACGAGAGTATTTATCCAAATATGAGAGGCTTTTACT	5356
QY	898	ATTGCTTGACGACGCTCTCTCGAAACAACTGGAATAGATTTACAGCTGTCAACATGCCA	957
Db	5357	ATTAAATGATCTCTGCTCAATTTGTTATTAATGATGAACTGATTAAGTCTGTATCTTACCA	5416
QY	958	GAAATCAGTGTCTCGGTTGTTGTTGCTCGGACGTTTGAATTTGATGCAAAAGTGGAA	1017
Db	5417	CAGATTAGTATTCAGTACTATGCTGCCACGACAACTTGACATTTATGTCAGAGCTAAAT	5476
QY	1018	AATCCCGAGGCTGACAAATATATGTTGTTAAATTAATGTTACTTTTACACCAAGTGTCAA	1077
Db	5477	AATCCGAGGCGGTGAGTATTCAGGAATGACTTAAGCTAACTTACTCTCTAGTAGTTCA	5536
QY	1078	ACACTCTA 1085	
Db	5537	AGTCTGTA 5544	
RESULT 10			
AX741415			
LOCUS	5798 bp	DNA	linear
DEFINITION	Sequence 5 from Patent WO03022306.		PAT 29-MAY-2003
ACCESSION	AX741415		
VERSION	AX741415.1		GI:30524208
KEYWORDS	Escherichia coli		
SOURCE	Escherichia coli		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1		
AUTHORS	Turner, A.K., Greenwood, J., Stephens, J.C., Beavis, J.C. and Darsley, M.J.		
TITLE	Bacterial vaccine		
JOURNAL	Patent: WO 03022306-A 5 20-MAR-2003;		
FEATURES	Acambis Research limited (GB)		
source	Location/Qualifiers		
	1..5798	/organism="Escherichia coli"	
		/col_type="genomic DNA"	
		/db_xref="taxon:562"	
	499..1215	/genes="cotB"	
gene	499..1215	/genes="cotB"	
CDS		/codon_start=1	

```
/transl_table=11
/protein_id="CAD90909.1"
/db_xref="GI:30524209"
/db_xref="REMTREMBL:CAD90909"
/translation="MKILLFVILFNVFAASANFWYPIKDIQSGSETIKVFSKSK
DVQYIKYIKRVINPQGTKEQVDIKWDGGLITVPKVLVPAAGSKSRLTINKE
QEVVVYFESVKPGQDDIEBKNGRVNLDLSNIIYAALIRTSPNPKLDVIES
NHWIKNTGNIRLGIKDVFLCDTTSINDKCAKFSYNNLYPDMSVITKLGKGFSAV
IDKDDNNSGGLINIKLP"
gene
1255..1767
/genes="cotA"
CDS
1255..1767
/genes="cotA"
/codon_start=1
/transl_table=11
/protein_id="CAD90910.1"
/db_xref="GI:30524210"
/db_xref="REMTREMBL:CAD90910"
/translation="VKLNKILGALVLSSTFVSMGASAAEKNIITVASVDPTIDMOSD
GTALPSAVNIATLPGKRFESARINQVHTNNYTKGLQIKLNDNVNINLSDPSKIL
PLEVSPAGTKLSTATSIITADQLNFGAAGVETVSTAKELVINAGTQQQINIVAGNYQ
LNSVILVTPP"
gene
1836..4436
/genes="cotC"
CDS
1836..4436
/genes="cotC"
/codon_start=1
/transl_table=11
/protein_id="CAD90911.1"
/db_xref="GI:30524211"
/db_xref="REMTREMBL:CAD90911"
/translation="MRAFNKITVFLFIPFLGRTNGLSKKNIPBEFIDLWQBDEL
LEVNLXGSLGHRVLTPTTKVPSVEEILEKINVKQKKDLRSLILQYSYRNGM
SCNGDFEXYSCNYRTDVTNVIDEENELNLFIGASFLSVAQNIYYQKINSEK
ATHTIDNFSSEGYKSLKSGVGAQGLTENSYLFGMDAIYNSRKTYTKQSNIN
IYRYPFDKYYQLGRMDRSLSSSGNFNMLPLPDIDGFOCTQSYTKNTEK
SISSPTVMTLTPESVEAFNEELLGVWVLSINDLDTSLPDGSVDLTLLI9EODI
LYREEKVPKKGASFGDMQWDFVFAQAGNIYNNDSYIEKQTKKKGGINAGITPVR
NUSFIGGAILINDKTYEAGVNRWSGLDGLSGNFSFLYGGDARGNYQNISTDGPN
LGFYRNDKDSCHNSYISAGWGCYSYSFSLSPVSGWTTTLGYNHTNNEAVHKDY
TPEYFYSKYKGVSKRWQTSYSSYKMDYHVIPTTGVVSDQSRSEQGYFSLST
RVKNSAINAGYSYVVKHKNATHAFLDGRITNTTGYSELGSRNTNKNNTAGVT
GAKVNEFDLNGSLNNVSKTSKMTSHMSANVNSFAITGDSVYWGDAAGLTKLGG
GVNVRSDDKRELKISGSSYGNVILGSDNRGFIPIVSLMPSNLIIEETQSDKNIT
VQALSKNDFIIPGNVPIDVTANVTVSIGRALDDKPLNSNAHLIDVHVRGLDSG
GFSFETSAQKSLFLKXDDYSCDVVKYDLRSGLVLTGDLTCEHSGIERLGLDLYNN
PRVKQLLAYK"
gene
4451..5545
/genes="cotD"
CDS
4451..5545
/genes="cotD"
/codon_start=1
/transl_table=11
/protein_id="CAD90912.1"
/db_xref="GI:30524212"
/db_xref="REMTREMBL:CAD90912"
/translation="MKRVIFVLSMFLCSOVYQSWTNEAGSINTKESIGPIDR3AA
ASYPAHYIPEHVAGVNGKHSFLDFRMTFLMGSTDSKACPTGSKSGSQGTINIKL
IFTEPKSLARKTINLNGYKRYFLYEDRCIHYDKMNLNGHTVKCVSGFTRGVDFLLYI
POGEIDGLTGGTWEATLELRVKRHYDYNHTYKVNITVDLTDKGNQVTPKFDSDP
RIDNLNRPNGKYSNVLEMLYDGYSTHSQSIEMRFDDSDQSGNNYNIKTGEP
LAKLYKLILLIGGREFVNGEAFVINDTSSLFINWNRKKSUSLPIQISIPVLCWPA
LTFMSELNPNPEAGYSGILNVTFTFSSSL"
```

ORIGIN

Query Match 27.5%; Score 298.8; DB 6; Length 5798;
Best Local Similarity 58.5%; Pred. No. 4.1e-54;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;
QY 127 GAATCTTCCCCCAACATAATATTAATAATACCATATTACACATACAGATGAAGTCAT 186
Db 4583 GCATCTGATCCCTCTCATATATATTTTCATGAACATGTTCTGTTACATAAAGATCAC 4542


```

1 none.
AUTHORS Genome sequence of the nematode C. elegans: a platform for
TITLE investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
AUTHORS 2 (bases 1 to 61864)
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1999) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England, and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jesse@sanger.ac.uk or rwenemate@wustl.edu
On Dec 15, 1999 this sequence version replaced gi:5832869.
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
[991122 dll] : Frameshift detected in sequence. Single nucleotide
removed from the project.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=y50E8A
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone Y50E8A.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone C48G7 is at 61761 in this sequence. The
true right end of clone R08A2 is at 106 in this sequence. The start
of this sequence (1..106) overlaps with the end of sequence Z98853.
The end of this sequence (61761..61864) overlaps with the start of
sequence Z78061.
FEATURES             Location/Qualifiers
     1..61864
     /organism="Caenorhabditis elegans"
     /mol_type="genomic DNA"
     /strain="Bristol K2"
     /db_xref="taxon:6239"
     /chromosome="V"
     /clone="Y50E8A"
     join(1923..2105,2179..2256)
     /gene="Y50E8A.1"
     join(1923..2105,2179..2256)
     /gene="Y50E8A.1"
     /standard_name="Y50E8A.1"
     /codon_start=1
     /product="Hypothetical protein Y50E8A.1"
     /protein_id="CAB5044.1"
     /db_xref="GI:5832870"
     /translation="WNFIITHILMSLSISIVTARPPGNDQEKLLTGGEPLGDDYV
DYPDFPVLNKSSTKLNNVQYKKYLSVAGFPAPKMKML"
     complement(3457..3756)
     /gene="Y50E8A.2"
     /gene="Y50E8A.2"
     /standard_name="Y50E8A.2"
     /codon_start=1
     /product="Hypothetical protein Y50E8A.2"
     /protein_id="CAB5045.1"
     /db_xref="GI:5832871"
     /translation="MRLRGVADPCGVLDPKETVLLDVSDAFAGQESNNDRLLTAVW
TNPNBAKQFRREGCGQGMTRKNSQLSTHLNKSVPFLVFFLSNKSFT"
     join(4673..4783,5619..5699,5756..5833,5909..6094)
     /gene="oig-3"

```

```

CDS

```

```

join(4673..4783,5619..5699,5756..5833,5909..6094)
/gene="oig-3"
/standard_name="Y50E8A.3"
/codon_start=1
/product="C. elegans OIG-3 protein (corresponding sequence
Y50E8A.3)"
/protein_id="CAB55046.1"
/db_xref="GI:5832872"
/translation="MSYSPFPQPQYQSSTLVNDCVGVRSQQVSPLTQVNHLY
IPKYNANPNWNAPEFLKVGDLRVHGVKAVDFCVLLGSPFRVCLWFNEDKLR
SDVIEDADVCLRLTIPVQVMEHFGCTFVLCNEVGRATAQAELLPLVE"
     complement(6492..9248)
     /gene="unc-61"
     complement(join(6492..6734,6791..7280,7600..7740,
8092..8173,8224..8363,8412..8506,8622..8795,8840..8893,
9069..9248))
     /gene="unc-61"
     /standard_name="Y50E8A.4a"
     /note="cDNA EST YK733g2.5 comes from this gene
cDNA EST YK733g2.3 comes from this gene"
     /codon_start=1
     /product="C. elegans UNC-61 protein (corresponding
sequence Y50E8A.4a)"
     /protein_id="CAD54172.1"
     /db_xref="GI:24817603"
     /translation="MSFTILITASALLFELVLITTAFLIVRSKQSSINLQTVV
THENPYVTWTSOSSKMDIPVILPKRFNMSDIEHKLPHQPPPPVPHHQNQPHNNT
TTISSATSINTTTSKXPTTAAPTAPSPKISLSDHTRGLMQLNGHVGDSPLQLVK
KAVAGAGFNLNMGVETGTGKTLIESLNKMLDEPCNHELKVELKTCCTDVAEGG
IRVKLRLVETAGFGDQDKSAKIVLYLSQFETYLQELKPMRLQYFNDSLHA
CLYFSPGHGLKALDLVTLRLAKRVNVIPIAKSDTTCDELLAKLSELSKQ
KIDITFFDDETVSTTKMKNKSVFPAVGSIDPVKENGQMVZARQYPMWIVVEN
ESHCDPVLKRLALTNVDEMRQTHESLYENRRDRKQMKIGSGTGPKLIEKLAQ
KHREQDEFRRRLTREFEQKLDVTGDMRKVEGLAREREREVHENYRREASKLDM
EIROTERRMKLMTKVSKLKLK"
     complement(join(6492..6734,6791..7280,7600..7740,
8092..8173,8224..8363,8412..8506,8622..8795,8840..8860))
     /gene="unc-61"
     /standard_name="Y50E8A.4b"
     /note="contains similarity to Pfam domain: PF00735 (Cell
division protein). Score=415.1, E-value=2.1e-121, N=1
cDNA EST YK109f1.5 comes from this gene
cDNA EST YK22f8.5 comes from this gene
cDNA EST YK411f2.3 comes from this gene
cDNA EST YK524f2.3 comes from this gene
cDNA EST YK411f2.5 comes from this gene
cDNA EST YK510f2.3 comes from this gene
cDNA EST YK510f2.5 comes from this gene"
     /codon_start=1
     /product="C. elegans UNC-61 protein (corresponding
sequence Y50E8A.4b)"
     /protein_id="CAB60583.1"
     /db_xref="GI:6425352"
     /translation="MSDIEHKLQHLPPHPPPPVPHHONQPHNNTTTSATSSIN
TTTTSKPTIAAPTAPSPKISLSDHTRGLMQLNGHVGDSPLQLKAVRAGVQPNL
MCVGTGKTKTLLIESLNKMLDFFCNHELKVELKTCCTDVAEGGIRVKLRLVETA
GFGDQDKSAKIVLYLSQFETYLQELKPMRLQYFNDSRLHACQYFISFGHG
LKALDLVTLRLAKRVNVIPIAKSDTTCDELLAKLSELSKQKIDITFFDDE
FTVSTTKMKNKSVFPAVGSIDPVKENGQMVZARQYPMWIVVENESHCDPVLKRE
ALLRNVDEMRQTHESLYENRRDRKQMKIGSGTGPKLIEKLAQKHREQDEFRR
RLTJUREBFOKLDVTGDMRKVEGLAREREREVHENYRREASKLDMELRQLTERVK
LMTYSKLRK"
     complement(join(9824..9902,9968..10102,10156..10258,
10304..10469,10683..10728,10840..10979,11047..11151))
     /gene="Y50E8A.5"
     complement(join(9824..9902,9968..10102,10156..10258,
10304..10469,10683..10728,10840..10979,11047..11151))
     /gene="Y50E8A.5"
     /standard_name="Y50E8A.5"
     /codon_start=1
     /product="Hypothetical p-rotein Y50E8A.5"
     /protein_id="CAB55048.1"
     /db_xref="GI:5832874"

```

```

gene
CDS

```



```
ELKSRVKLNVRIRPTDCKKJGGVLOIHEINGRLVYOSQPOSQSNFKNDQVDV"
PSNIXHLPPOCKDELIVLHCHLKNPIMICKRKTEDVOFYEA5DMAFDETGRKERK
IRYGDDELQEQBERRRKALJDEKFKGFA5LADSHGWLDLIPFRELFGQVPPFR
SSVLCPRDCLVQIDIPYLVITLSELEIAHLERVOFGLNFDLVPFKDFNKPVEH
INTIPVELLEDKWLVDVPIBEGQNLNWKVIMKTIVLADPQCFIDGGAFLITGQ
GSEDEESDESDFRVSDEDPQDEDEESDDYASGEESDDYSGDDDDGGGGDDDDSE
SGEDVALERKAADNRSGFD"
/sequence=<4198..>5230)
/gene="cso99"
/complement (<4198..>5230)
/gene="cso99"
/complement (<4198..>5230)
/product="Cso1p"
/complement (<4198..>5230)
/gene="cso99"
/codon_start=2
/transl_table=12
/protein_id="Cso1p"
/db_xref="GI:12082671"
/translation="IKYFNITTTQPIIMSSISGISNIISSIKNGSNKPTLLPISI
LLNLOKLKKSIFSKFQOEINDKYSKSSVDQNLVFCYPSLLISAILNKKKILY
FLNKWYFIOLVKSATGISTTTATATTTATTKKEENNGSTSVLSKILPKPVPYT
EGTPKLAH:KKINSYADIRIFNRLTDSIKYMEWIDYEHAFNTNSPTSIKPRF
FINLLQSNCLVLELLENLWTDHWTGTGNNEWCTETWCCRVGVLIIIME
IIRTPMTPKWSNNKQWNLNPKNVQLPLVLHWSLRDGLTPFWVGLCGSGASWNEFX
DNWSSIDLS"
ORIGIN
Query Match 4.9%; Score 53.6; DB 8; Length 5230;
Best Local Similarity 47.8%; Pred. No. 0.2;
Matches 155; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 733 ACTACAGAGCTCTTGGAGATAGATTTCAGGATACAACTCCAAATCTCATGGGAA 792
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 733 TTTTATCTAAGCAAAATAATGATGACACCAAGAAATGCAATATCTTGTCACTTCTC 852
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4946 TTTAAATGTCGAATTAATTAATGAATAATGAATAATGAATAATGAATAATGA 5005
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 853 TTGGCGGGTAAAGTTTAACTCCACAAATGAAAGCTCATTTAAATTTGCTGACGAGCT 912
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5006 GTTGAAGATTACTTAATCTTATCATTAATTTCTTGATAAAATTTCTTGAAAAAAT 5065
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 913 TCTCTGGAACAACTGGAATAGATTACAGCTGTCACCATGCCGAATAGTGTTCG 972
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5066 GATTTTTCAAAATTTTTCAGAGATTTAATAATATAGATATTGGTAATATGTTTAGGT 5125
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 973 GTGTTGTGTGGCTGGAGCTTTGCAATTTGGATGCAAAAGTGGAAATCCCGAGGCTGA 1032
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5126 TTATTGGAATTAACATTGAGTTTGAATTTGAATTTGAATTTGAATTTCCGAAATTGA 5185
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1033 CAATATATGGGTAACATTAACTTT 1056
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5186 CTCAATTATCGTTGTGTTGTTGTT 5209
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 14
AC115594
LOCUS Dictyostelium discoideum chromosome 2 map 4071862-4101005 strain
DEFINITION AX4 complete sequence.
AC115594
VERSION AC115594.2 GI:28828945
KEYWORDS Dictyostelium discoideum
SOURCE Dictyostelium discoideum
ORGANISM Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 29143)
AUTHORS Gloeckner, G., Eichinger, I., Szafarski, K., Pachebat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., April, J.F., Guigo, R., Rumpf, K.,
Tunggal, B., Cox, E., Quail, M.A., Platzer, M., Rosenthal, A. and
Noegel, A.A.
```

```
TITLE
JOURNAL Nature 418 (6893), 79-85 (2002)
MEDLINE 22092822
PUBMED 12097910
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 29143)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
REFERENCE 3 (bases 1 to 29143)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
REFERENCE 4 (bases 1 to 29143)
AUTHORS Baumgart, C.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
COMMENT On Mar 4, 2003 this sequence version replaced gi:19570001..
CDS predictions from Genbank do not necessarily reflect true genes.
Further information is available from IMB Jena, Department of
Genome Analysis
'http://genome.imb-jena.de/dictyostelium/'
and the University Cologne, Institute for Biochemistry I
'http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
Location/Qualifiers
1. .29143
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="4071862-4101005"
/join("<769..1194,1286..2293)
/notes="GeneID exon scores (in order of location ranges):
39.90, 122.92 - GSCJ_ID dd_00581"
/codon_start=1
/product="similar to Fundulus heteroclitus (Killifish)
(Murichog). Cytochrome P450 2N1"
/protein_id="AA051538.1"
/db_xref="GI:28828958"
/translation="FHKRTGMPGPPPLPIGLHQLGKSPYKSLKSFSDKYGGLT
TIFGSVPTVLISENLRLRIIKNDSIIDVYSDGLIIGGERNLLFSGKSPWKY
RKIFSSAMTARKFNIAIRIQQAISNNYFTYANSKOAINPHDYIRYSLNGVIDY
SPFSVYESDTHIVIRAAEMIEILATGNPDYLPFLPPYTKRNTLAVAGVW
DYCDAITVHRKTLDHKEPRDLIDLIMEIKSEKQFVDDSLSKCLTDLVAGHET
VAITLGMILLFLSNHQYQVYDELINVWKGKGLPALVHRKDTSLNACIQTETIR
TAAPLAPRIASEDKVGYTIPKGTQNMVYGMASDERYWKDQSHIFNPSRLSNR
STENGCGGGVGVNSSQSEVPIFGVGRMCMVGMAKDELYYCQSMFMPNPKWSPVN
DKPIDEGVARIATLEYQVYVLEK"
/join(2894..3032,3129..3249,3358..4105)
/notes="GeneID exon scores (in order of location ranges):
9.13, 15.24, 82.27 - GSCJ_ID dd_03366"
/codon_start=1
/product="similar to Ralstonia solanacearum (Pseudomonas
solanacearum). Probable alcohol dehydrogenase (ADH-HT)
oxidoreductase protein (EC 1.1.1.1)"
/protein_id="AA051530.1"
/db_xref="GI:28828946"
/translation="MKAAVFKQKGDLSIVEVPPEPKQVVRKVKHSCGVCHSDNAC
RYGNFNSYPRIPGHEVGEIKLGEVDPNYFKICFVGVGFGHCGKRCLEND
QWIFCKSLICGIIHYDGGVAYBMVAPSDSLAFVPEGMAAETATVFCAGITVFNFRN
QNIKVGSLVGVQGLGHLAIQFCKMKGVYVIAMSGDQKELARELCAAHYVNIK
DGWIDOMEIKSQVCIILTPASSIVQSSLDALGVNGKLVLAILEPENAHSYILIG
GNKSTIGSSSDSRDAECTLPANKNQKAMKTFFLENVNEALLKIDARPHRVKI
1"
/join(5140..5233,5445..5565,5640..6387)
/notes="GeneID exon scores (in order of location ranges):
3.85, 11.61, 58.39 - GSCJ_ID dd_00585"
CDS
```

```
FEATURES
source
```

```
FEATURES
source
```


Search completed: May 8, 2004, 10:56:04
Job time : 4400 secs

(G alpha 2),
 /protein_id="AA052223.1"
 /db_xref="GI:28829710"
 /translation="MGICASSWGEFKEFKDILIRNLDNMRVLDDGRLGTMITDPSNDRAY
 STISKQMKIIHOSGYNIEKREFKDILIRNLDNMRVLDDGRLGTMITDPSNDRAY
 MIKELTSLQASATPTDQWELNEEDQCKIKALWTFDGVKQAMRRANERSTLPDGAPEF
 DEEDWTSFVPTPTDQWELNEEDQCKIKALWTFDGVKQAMRRANERSTLPDGAPEF
 FQDVAIVFCVALSYDILLVEDNSTNEMLSLVSFSDVCSWPNPILILFLNKSPL
 FREKIKHVDLSSTFEYKGGRIYERASNIINERFWQINKTEQKALISHICATDTNNI
 RVFPEAVKDIIIFTCQWKAAGLS"
 join(122686..22743,22820..22926,23004..23139,23232..23558;
 23639..27180)
 /note="GeneID exon scores (in order of location ranges):
 -4.93, 3.57, -1.17, 18.69, 174.64 - GSCJ_ID dd_01857"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA052224.1"
 /db_xref="GI:28829711"
 /translation="MNRNHLHVFVLSILSLINGIKSVKVGDFSVIDIITDSFDDLV
 TLNVDVSGGLDENIFWIKVNTVOIGENQKNLLTLDNSTFTIQPDFDTLACKER
 PVNIRELIVQCKFDKDYKQVGLHVNIGLEOYLSVPGLENNFISNTFGQYDDDS
 IFYDYSKNTTFIKPVIKASAVISQLSIDTFGVSSTSKLYLTFSLFNSFYKNPIT
 APLQNDNCSLLVDCSLFVQADFQPGHIFLNMNDSNLPELVGNKSGVHWL
 FTKNAERGGSYIYESFSDTFGTFISNIFSGFTGVNVLVYI00SFYGCIQN
 SFLFSGFVPTPNVYGIDFIDNGENTLSHPKIFPNTINAGCAANTLWVPSPPY
 QKSLAEITLQFPIKPSGILVGGISGIAFKESGIPVFNIEYFSDSDCVIRIS
 ASDEGSCQSIASLPIGAVESPQVSIIRTNRLIEGDCPKGIYQISIPKYSIRRN
 YFEMXDYANNLKVLPNNFYSGKRYLGSPPYCVLTLEDIESIGPRNNNIDLSNNS
 NVLYIKKQISNLVQFIPKIGITDYTPNNKISGTPRFPKSKWNSASSIYEVSEDI
 PASLFGPISYILNLEIDSSILQYFSGSENLVTSQIADNFGALSYNTIGNTLIT
 PSINPEPTFOILFDGLNGLSQVNTITSDTKYQKMFNTFIESLGRWQFTW
 EIKYGIPTFISLSDNSGYITFEPNPKVAPMLLVELYSPFSIQSGEPTPD
 VTPYQVHVNQLNENLVDRLLEINFVYDCEGSIIEEPTPIYFINDVEYIISKQ
 TSVLSGTSGLGFENAFQVNTIPYLVGPNGLTQHGIVDNKFNVLGATPIDIN
 QYNRGNSFNVITGQSQPLESFMVDIDGILLTVNCLKEGILPTCKLRINGALENLI
 SLSFLANTYIVFLKQTRNINDNCQIYITNQLGFSNNEINFLVKAHNVIMPTPTTP
 HPTSNPATSLFANVTPPLPSGPKCKDGGIAQGCQCNLCGICSPYIGIDCSSK
 VIVVIDIVNNSQFVITPYQHSNKGFTSVISVLSRELFGDKVENDYIFENWY
 QYNDNNTYFTGICKNKDLSTKTIKAYKQLTITQFAGENITQPESIKYFIS
 LSVNQRNSLNTLIDLVNLTETKEKIDDAFCSEFGDSQSLSQVQLKIQINDVSLG
 EPTKRAIVDTHVGVLNKVGKNNINSTNGSSSSQVSIQIPNYKTSVELDENFVLL
 KSKENLSTCNKQSGILTKTQLIGTAGSFILILIVISVIYKSKSTLLEINRA
 KRFKKKLSFK"
 complement(join(27340..29056,29130..29490))
 /note="GeneID exon scores (in order of location ranges):
 218.38, 14.76 - GSCJ_ID dd_01858"
 /codon_start=1
 /product="similar to Rhizobium loti (Mesorhizobium loti).
 Dehydrogenase, succinatesemialdehyde dehydrogenase,
 4.7%; Score 50.6; DB 3; Length 88549;
 Best Local Similarity 49.8%; Pred. No. 0.59;
 Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
 Query Match
 360 TAACAATTAATGTTCAAAGTGTTAACTGCCATCCGGCTAACACTTAACCTCAGCTCA 419
 Db 35734 TACCACATTAATTAATGAATGATGGATTCACTTCATTTCAACCACTTGATGAATC 35793
 QY 420 TTTTAACTGTATAAAAAAGCGGGCTTCAGGTGCAAGTTTATATTTATATATTCCTGCTGG 479
 Db 35794 TTTCATATTCAAGAGAATCAACAATATGGAAGCTAAGATTTAAATATTTATTATGG 35853
 QY 480 CGAAGTAAAAATTTGCCCTTTTGGTGGTATCTGGGATGCTACTCTGAGCTAAGCTAAGTAA 539
 Db 35854 TGAATATTATATGTAATATGATATTAAATATTCAAGATTATATATATCATATAAAATTAAT 35913
 QY 540 AAGACCATATAGTGAGACTATGGAAGCTATACACATATAAATCAATCAATTAATTAATTAATCA 599
 Db 35914 ATTACCAATTTAAGATCAATCAATTAATAATAATAATAATAATAATAATAATAATA 35973
 QY 600 TAAGGGAATATTTCAGA 616
 Db 35974 TAATAATAATAATAATA 35990

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 08:09:10 ; Search time 491 Seconds
(without alignments)
9396.218 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: 1 atgaataagattttattat.....caagtagtcaaacactotag 1086

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 segs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002s:*
- 7: Geneseq2003as:*
- 8: Geneseq2003bs:*
- 9: Geneseq2003cs:*
- 10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	6	AAI70763
2	1086	100.0	1086	8	ABX93757 DNA encod
3	1086	100.0	7239	6	AAI70780 ETEC CS4
4	1086	100.0	7239	7	AD55748 Escherich
5	1086	100.0	7239	8	AD55774 Csa opero
6	310.8	28.6	5336	7	AD55743 Escherich
7	298.8	27.5	5798	2	AAx84848 CS2 gene
8	298.8	27.5	5798	7	AD55745 Escherich
9	50	4.6	2000	7	ADA71938 Rice gene
10	49.4	4.5	10279	6	ABL33591 Human imm
11	49.4	4.5	10279	6	ABL92277 Chemicall
12	49.4	4.5	10279	6	AD22328 Chemicall
13	48.2	4.4	9504	4	ABK28407 DNA trans
14	47.2	4.3	5139	3	AAa70139 Plasmodiu
15	45.2	4.2	4176	4	ABL10866 Drosophil
16	44.8	4.1	13133	6	ABK31231 Signal tr
17	44	4.1	2000	7	ADA71938 Rice gene
18	43.8	4.0	6971	6	ABN80061 Human che
19	43.4	4.0	2036	4	ABL01847 Drosophil
20	43.4	4.0	6080	4	ABL01846 Escherich
21	43.2	4.0	50000	6	ABL55643 AmbEV gen
22	42.8	3.9	4985	6	ABq75107 Anopheles
23	42.8	3.9	4985	9	ACf79720 Mosquito

C 24	42.8	3.9	6137	6	ABL34450 Human met
C 25	42.8	3.9	6137	6	ABE70127 Chemicall
C 26	42.6	3.9	641	6	ABQ56894 Human col
C 27	42.4	3.9	19734	6	ABL33933 Human imm
C 28	42.2	3.9	9095	6	ABK28448 DNA trans
C 29	41.6	3.8	6209	4	AAx284751 Genomic s
C 30	41.6	3.8	15548	6	ABL34155 Human imm
C 31	41.6	3.8	110000	6	ABA90521_10 Contamination (11 o
C 32	41.2	3.8	454	7	ABT40601 Toxicity
C 33	41.2	3.8	454	7	ABT40601 Toxicity
C 34	41.2	3.8	16228	6	ABT40601 Toxicity
C 35	41.2	3.8	16228	6	ABT40601 Toxicity
C 36	41.2	3.8	16228	6	ABT40601 Toxicity
C 37	41.2	3.8	16228	6	ABT40601 Toxicity
C 38	41.2	3.8	16228	6	ABT40601 Toxicity
C 39	41.2	3.8	16228	6	ABT40601 Toxicity
C 40	41.2	3.8	16228	6	ABT40601 Toxicity
C 41	41.2	3.8	16228	6	ABT40601 Toxicity
C 42	41.2	3.8	16228	6	ABT40601 Toxicity
C 43	40.8	3.8	6534	6	ABL32470 Human imm
C 44	40.8	3.8	7833	9	ADB54183 Pretreate
C 45	40.8	3.8	7833	9	ADB54183 Pretreate

ALIGNMENTS

RESULT 1
AAI70763
ID AAI70763 standard; DNA; 1086 BP.
XX AAI70763;
AC AAI70763;
XX
DT 18-FEB-2002 (first entry)
XX
DE ETEC CS4 pilus csae gene coding region.
XX
KW CS4 pilus; enterotoxigenic; ETEC; csae operon; csae gene; fimbrial;
KW vaccine; diarrhoea; antibacterial; antidiarrheic; cs.
XX
OS Escherichia coli.
XX
EH Key Location/Qualifiers
FT s-g_peptide i..69
FT /*tag= a
FT mat_peptide 70..1083
FT /*tag= b
XX
XX WO200181582-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US012914.
XX
XX 20-APR-2000; 2000US-0198686P.
XX
XX (JYMA-) UNIV MARYLAND BALTIMORE.
XX
XX Altboun Z, Levine MM, Barry EM;
XX
XX WPI; 2002-049280/06.
XX
XX P-PSDB; AAM50343.
XX
XX New nucleotide sequence, useful as immunogenic agent for generating
XX immune response against recombinant product of the operon, comprises csae
XX operon which encodes enterotoxigenic Escherichia coli-CS4 pilus.
XX
XX Claim 31; Page 57-58; 81pp; English.
XX
XX The present sequence is that of the coding region of the csae gene of
XX enterotoxigenic Escherichia coli (ETEC) strain E11881A. The csae operon
XX (see AAI70780) was isolated from a genomic library of this strain.
XX
XX Sequencing revealed 5 contiguous genes, csae-csaE (see AAI70759-63),

CC flanked by 2 insertion elements. The *csa* operon encodes the synthesis of
 CC ETEC-CS4 pili, and has been expressed in attenuated Shigella strain
 CC CVD1204 guABA, constructing the Shigella expressing CS4 fimbriae vaccine
 CC strain CVD1204 (pGA2-CS4). The *csaE* gene encodes a 40 kDa tip associated
 CC protein (CS4D, see AM50341) that shows homology with similar proteins
 CC from other ETEC fimbriae. The *csa* operon, and the *csaA-csaE* genes, are
 CC useful in the production of recombinant CsaA-CsaE polypeptides that are
 CC used in claimed immunogenic compositions to prevent ETEC colonisation,
 CC and hence to protect against diarrhoea
 XX
 SQ Sequence 1086 BP; 366 A; 187 C; 202 G; 331 T; 0 U; 0 Other;
 Query March 100.0%; Score 1086; DB 6; Length 1086;
 Best Local Similarity 100.0%; Pred. No. 6 5e-254;
 Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGAATAAGATTTTAAATTTTACATTTGTTTCTTCTCAGTACTTTTACATTTGCT 60
 DB 1 ATGAATAAGATTTTAAATTTTACATTTGTTTCTTCTCAGTACTTTTACATTTGCT 60
 QY 61 GTATCGGAGATAAAATCCCGGAGATCAAGCATACTAATATTTTGGCCCGGTGAC 120
 DB 61 GTATCGGAGATAAAATCCCGGAGATCAAGCATACTAATATTTTGGCCCGGTGAC 120
 QY 121 AGGAACGAATCTCCGCCAAACATAATATTAATAAACCATATTACAGCATACAGTGAA 180
 DB 121 AGGAACGAATCTCCGCCAAACATAATATTAATAAACCATATTACAGCATACAGTGAA 180
 QY 181 AGTCATATCTGTATGATAGATGACATTTTATATGTTTGTCTTCTCACAATACACTTAAT 240
 DB 181 AGTCATATCTGTATGATAGATGACATTTTATGTTTGTCTTCTCACAATACACTTAAT 240
 QY 241 GGACATGTCCAAACAGTCAGAACTCTAGCAGTTCATCGGTGCGGTGAAACAAATATA 300
 DB 241 GGACATGTCCAAACAGTCAGAACTCTAGCAGTTCATCGGTGCGGTGAAACAAATATA 300
 QY 301 ACATTACAATTTACGAAAAAAGAAAGTTTAAATAAAAGAGAGCTACAAATTAAGGCTAT 360
 DB 301 ACATTACAATTTACGAAAAAAGAAAGTTTAAATAAAAGAGAGCTACAAATTAAGGCTAT 360
 QY 361 AAACAATTTATTTTCAAAGTGTAACTGCGCCATACCTTAACTTAACTCAGTCAAT 420
 DB 361 AAACAATTTATTTTCAAAGTGTAACTGCGCCATACCTTAACTTAACTCAGTCAAT 420
 QY 421 TTTTAAGTGAATAAAACGGCGCTTACAGTCAAGTTCATATTTATATTTATTTTCTGGC 480
 DB 421 TTTTAAGTGAATAAAACGGCGCTTACAGTCAAGTTCATATTTATATTTATTTTCTGGC 480
 QY 481 GAACTAAAAAATTTGCGCTTTTGGTGTATCTGGATGCTACTCTGAAGTTAAGAGTAAAA 540
 DB 481 GAACTAAAAAATTTGCGCTTTTGGTGTATCTGGATGCTACTCTGAAGTTAAGAGTAAAA 540
 QY 541 AGACGATATAGTGAACCTATGGAACTTACACTATAAATATCACTATTAATTAACGTAT 600
 DB 541 AGACGATATAGTGAACCTATGGAACTTACACTATAAATATCACTATTAATTAACGTAT 600
 QY 601 AAGCGAATATTCAGATATGTTTCTCAGTTTCAAAAGTGCCTGCGTCTGATTAAC 660
 DB 601 AAGCGAATATTCAGATATGTTTCTCAGTTTCAAAAGTGCCTGCGTCTGATTAAC 660
 QY 661 TTGCTTCAAACTGTTGGGGGACATATATTTGGAAGAAATTTCTTGAATATGCTTTTAT 720
 DB 661 TTGCTTCAAACTGTTGGGGGACATATATTTGGAAGAAATTTCTTGAATATGCTTTTAT 720
 QY 721 GATGATATAGTAACTAAACAGAGCTCTTTGGAGATAAGATTTTCAGGATAACATTCCTAAA 780
 DB 721 GATGATATAGTAACTAAACAGAGCTCTTTGGAGATAAGATTTTCAGGATAACATTCCTAAA 780
 QY 781 TCTGTATGGAATTTTATCTAAGGAATAATAATGATGACACCAAGAAATTCATATACT 840
 DB 781 TCTGTATGGAATTTTATCTAAGGAATAATAATGATGACACCAAGAAATTCATATACT 840
 QY 841 TTGTGACCTCTCTTGGCGGTAAAAAGTTTAACTCCAAACAAATGGAACGTCATTAATAT 900

DB 841 TTGTCACTTCTCTTGGCGGTAAAAAGCTTTAACTCCAAACAAATGGAACGTCATTAATAT 900
 QY 901 GCTGACGGAGCTTCTCTGGAACCAACTGGAATAGATTACAGCTGTCAACCATGCCAGAA 960
 DB 901 GCTGACGGAGCTTCTCTGGAACCAACTGGAATAGATTACAGCTGTCAACCATGCCAGAA 960
 QY 961 ATCAGATGTTCCGGTGTGTTGTTGGCTTGGACGTTTGCATTTGGATGCAAAAGTGGAAAT 1020
 DB 961 ATCAGATGTTCCGGTGTGTTGTTGGCTTGGACGTTTGCATTTGGATGCAAAAGTGGAAAT 1020
 QY 1021 CCGAGGCTGGACAATATATATGCGTAAATATTAATGTTACTTTTCACCAAGTACTCAACA 1080
 DB 1021 CCGAGGCTGGACAATATATATGCGTAAATATTAATGTTACTTTTCACCAAGTACTCAACA 1080
 QY 1081 CTCTAG 1086
 DB 1081 CTCTAG 1086
 RESULT 2
 ABX93757
 ID ABX93757 standard; DNR; 1086 BP.
 XX
 XX ABX93757;
 XX AC
 XX
 DT 16-JUN-2003 (first entry)
 XX
 DE DNA encoding *csa* operon recombinant product, CsaE.
 XX
 XX CsaE; gene; ds; *csa* operon; CS4 pilus; bacterial pili protein; ETEC; CS4;
 KW enterotoxigenic Escherichia coli; multivalent Shigella-ETEC; diarrhoea;
 KW Shigella; antibacterial.
 XX
 XX Escherichia coli.
 XX
 FH Key Location/Qualifiers
 CDS 1..1086
 FT /*tag= a
 FT /product= "CsaE"
 FT
 XX
 FN US2002:76868-A1.
 XX
 PD 28-NOV-2002.
 XX
 XX 20-APR-2001; 2001US-00839894.
 BP
 XX
 XX 20-APR-2000; 2000US-0198626P.
 PR
 XX (ALTB/) ALTBOUY Z.
 PA (LEVI/) LEVINE M M.
 PA (BARR/) BARRY E M.
 XX
 XX Altbouy Z, Levine MM, Barry EM;
 FI
 XX
 DR WPI; 2003-352604/33.
 XX P-PSDB; ABU09014.
 DR
 XX
 DR Novel isolated nucleic acid comprising *csa* operon encoding proteins
 PT required for producing CS4 pili, useful for generating immune response in
 PT vertebrate against the enterotoxigenic Escherichia coli.
 XX
 XX
 PS Claim 31; Page 30-31; 58pp; English.
 XX
 CC The invention relates to an isolated nucleotide sequence comprising a *csa*
 CC operon (encoding proteins required for producing CS4 pili) or its
 CC functional fragment. An immunogenic composition comprising a recombinant
 CC product of a *csa* operon and a carrier, is useful for generating an immune
 CC response in a subject, which involves contacting the subject with the
 CC immunogenic composition. The recombinant product of the composition is
 CC the CS4 antigen (bacterial pili protein) and is provided in an acellular
 CC or cellular composition. The nucleic acid is useful for producing a
 CC polypeptide product from a *csa* operon or functional fragment, which

involves providing a nucleic acid, introducing the *csa* operon in an expression vector, such that a recombinant host cell is produced and expressing the recombinant host cell to conditions such that a protein from the *csa* operon is expressed. The nucleic acid encoding at least an immunogenic portion of the *csa* operon or a polypeptide encoded by the nucleic acid is useful for generating an immune response in a vertebrate against ETEC (enterotoxigenic *Escherichia coli*). The nucleic acid is also useful for identifying polynucleotides encoding other proteins with biological functions similar to that of the *csa* operon and for creating a multivalent Shigella-ETEC immunogenic composition that will protect from diarrhoea caused by either Shigella or CS4 expressing ETEC strains. This sequence represents DNA encoding a recombinant product of a *csa* operon, CsaE

[illegible]

Query Match	Score 1086;	DB 8;	Length 1086;
100.0%			

Query Match	100.0%;	Score 1086;	Length 254;
Best Local Similarity	100.0%;	Pred. No. 6.5e-254;	
Matches 1086:	Conservative	0;	Mismatches 0;
	Indels	0;	Gaps 0;

Best Local Similarity 100.0%; Pred. No. 6.3e-254;
Matches 1086: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

Db	781	TTCTGATGGGAAATTTTATCTTAAGGAAATAAATGATGACACCAAGAAATTCATATACT	840
QY	841	TTGTCACTTCTCTTGGCGGCTAAAAGTTTAACTCCAAAGAAATGGACGTCATTAAATAT	900
Db	841	TTGTCACTTCTCTTGGCGGCTAAAAGTTTAACTCCAAAGAAATGGACGTCATTAAATAT	900
QY	901	GCTGACGAGCTTCTCTGGAAACAAACTGGAATAGAAATTACAGCTGTCCACCATGCCAGAA	960
Db	901	GCTGACGAGCTTCTCTGGAAACAAACTGGAATAGAAATTACAGCTGTCCACCATGCCAGAA	960
QY	961	ATCAGTGTCCGGGTGTGTGTGGCTGGACGTTTGCAATTGGATCAAAAGTGGAAAT	1020
Db	961	ATCAGTGTCCGGGTGTGTGTGGCTGGACGTTTGCAATTGGATCAAAAGTGGAAAT	1020
QY	1021	CCCGAGCGTGGCAATATATGGTGAATATAATAGTTACTTTTCAACCAAGTAGTCAACA	1080
Db	1021	CCCGAGCGTGGCAATATATGGTGAATATAATAGTTACTTTTCAACCAAGTAGTCAACA	1080
QY	1081	CTCTAG 1086	
Db	1081	CTCTAG 1086	
RESULT 3			
AAI70780			
XX	ID	AAI70780 standard; DNA; 7239 BP.	
XX	AC	AAI70780;	
XX	AC		
XX	DT	18-FEB-2002 (first entry)	
XX	XX		
DE	DE	ETEC C54 pilus csa operon.	
XX	XX		
KW	KW	C54 pilus; enterotoxigenic; ETEC; csa operon; csaA gene; csaB gene;	
KW	KW	csaC gene; csaD gene; csaE gene; csaF gene; fimbrial; vaccine; diarrhoea;	
KW	KW	antibacterial; antidiarrheic; ds.	
XX	XX		
OS	OS	Escherichia coli.	

```

XX
FH      Location/Qualifiers
CDS     283..999
        /*tag= a
        /note= "csaA (claim 15), encodes AAM50339"
FT
FT
FT      sig_peptide
        283..342
        /*tag= b
FT
FT      mat_peptide
        343..996
        /*tag= c
FT
FT      CDS
        1028..1531
        /*tag= d
FT
FT      /*note= "csaB (claim 22), encodes AAM50340"
FT
FT      sig_peptide
        1028..1096
        /*tag= e
FT
FT      mat_peptide
        1097..1528
        /*tag= f
FT
FT      CDS
        1589..4192
        /*tag= g
FT
FT      /*note= "csaC (claim 25), encodes AAM50341"
FT
FT      sig_peptide
        1589..1654
        /*tag= h
FT
FT      mat_peptide
        1655..4189
        /*tag= i
FT
FT      CDS
        4196..5391
        /*tag= j
FT
FT      /*note= "csaE (claim 31), encodes AAM50343"
FT
FT      sig_peptide
        4196..4364
        /*tag= k
FT
FT      mat_peptide
        4265..5278
        /*tag= l
FT
FT      CDS
        complement=(5790..6119)
        /*tag= m
FT
FT      /*note= "(claim 28) csad, encodes AAM50342"
FT
XX

```

PN WO200181582-A2.
 XX
 PD
 XX 01-NOV-2001.
 PF 20-APR-2001; 2001WO-US021214.
 XX 20-APR-2000; 2000US-0198686P.
 XX (UYMA-) UNIV MARYLAND BALTIMORE.
 PA
 XX
 PI Altboum Z, Levine MX, Barry EM;
 XX
 XX WPI; 2002-049280/06.
 DR P-PSDB; AAM50339, AAM50340, AAM50341, AAM50342, AAM50343.
 XX
 DR New nucleotide sequence, useful as immunogenic agent for generating
 XX immune response against recombinant product of the operon, comprises *csa*
 PT operon which encodes enterotoxigenic *Escherichia coli*-CS4 pili.
 PT
 XX
 XX Disclosure; Page 63-70; 81pp; English.
 XX
 CC The present sequence is that of the *csa* operon of enterotoxigenic
 CC *Escherichia coli* (EPEC) strain E-11891A. The *csa* operon was isolated from
 CC a genomic library of this strain. Sequencing revealed 5 contiguous genes,
 CC *csaA*-*csaE*, the coding regions of which are claimed (see AAL70759-63),
 CC flanked by 2 insertion elements. The *csa* operon encodes 5 proteins (see
 CC AAM50339-43) involved in the synthesis of EPEC-CS4 pili. These are the
 CC fimbrial structural protein (*CsaB*), the top associated protein (*CsaE*), a
 CC chaperon-like protein (*CsaA*), an usher-like protein (*CsaC*), and a
 CC truncated regulatory protein (*CsaD*). The *csa* operon has been expressed in
 CC attenuated *Shigella* strain CVD1204 guaBA, constructing the *Shigella*
 CC expressing *CS4* fimbriae vaccine strain CVD1204 (pGA2-CS4). The *csa* operon
 CC is useful in the production of recombinant *CsaA*-*CsaE* polypeptides that
 CC are used in claimed immunogenic compositions to prevent EPEC
 CC colonisation, and hence to protect against diarrhoea
 XX
 SQ Sequence 7239 BP; 2436 A; 1181 C; 1345 G; 2277 T; 0 U; 0 Other;

Query Match 100.0%; Score 1086; DB 6; Length 7239;
 Best Local Similarity 100.0%; Pred. No. 1.le-253;
 Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATAAGATTATTTATTTTACATGTTTCTTCCTTCAGTACTTTTACATTTGCT 60
 DB 4196 ATGAATAAGATTATTTATTTTACATGTTTCTTCCTTCAGTACTTTTACATTTGCT 4255

QY 61 GTATCGGCAGATAAAATTCGGAGATGAAAGCATATTAATTTTGGCCCGGTGAC 120
 DB 4256 GTATCGGCAGATAAAATTCGGAGATGAAAGCATATTAATTTTGGCCCGGTGAC 4315

QY 121 AGGAACGAATCTTCCCAACATATATATTAATAACCATATACAGCATACAGTGAA 180
 DB 4316 AGGAACGAATCTTCCCAACATATATATTAATAACCATATACAGCATACAGTGAA 4375

QY 151 AGTCATACCTGTATGATGAGGATGCTTTTATTTTCTTCCTTCAGTACTTTTACATTTAT 240
 DB 4376 AGTCATACCTGTATGATGAGGATGCTTTTATTTTCTTCCTTCAGTACTTTTACATTTAT 4435

QY 241 GGAGCATGCTCCCAACGATGAGATCCTAGCATTCATCGTCAGCGGTGAACAATATA 300
 DB 4436 GGAGCATGCTCCCAACGATGAGATCCTAGCATTCATCGTCAGCGGTGAACAATATA 4495

QY 301 ACATATACATTTTACGGAAAAAGAGTTTAAATAAAGAGAGCTACAAATTAAGGCTAT 360
 DB 4496 ACATATACATTTTACGGAAAAAGAGTTTAAATAAAGAGAGCTACAAATTAAGGCTAT 4555

QY 361 AAACAATTTATTTTCAAAAGTTTAACTCCGCGATTAACACTTAACTCAGCTCAT 420
 DB 4556 AAACAATTTATTTTCAAAAGTTTAACTCCGCGATTAACACTTAACTCAGCTCAT 4615

QY 421 TTTAAGTGAATAAATAACCGGCTTCAGTGAAGTTTATATTTATATTTATTTTCTGCTGC 480
 DB 4616 TTTAAGTGAATAAATAACCGGCTTCAGTGAAGTTTATATTTATATTTTCTGCTGC 4675

QY 461 GAACATAAAATTTGCTTTTGGTGGTATCTGGATGCTACTCTGAGTTAAGAGTAAAA 540
 DB 4676 GAACATAAAATTTGCTTTTGGTGGTATCTGGATGCTACTCTGAGTTAAGAGTAAAA 4735

QY 541 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATATCACTATTAAATTAACCTGAT 600
 DB 4736 AGACGATATAGTGAGACCTATGGAACCTTACACTATAAATATCACTATTAAATTAACCTGAT 4795

QY 601 AAGGAAATATTCAGATATGTTTACCTCAGTTCAAAAGTGACCTCGCTGCGATCTTAAC 660
 DB 4796 AAGGAAATATTCAGATATGTTTACCTCAGTTCAAAAGTGACCTCGCTGCGATCTTAAC 4855

QY 661 TTGCTCCAACTGGTGGGGCAATATATTGGAGAAATTTCTGTTGATATGCTGTTTAT 720
 DB 4856 TTGCTCCAACTGGTGGGGCAATATATTGGAGAAATTTCTGTTGATATGCTGTTTAT 4915

QY 721 GATGGATATAGTACTTAACAGCAGCTCTTTGAGATAGATTTCAGGATAACAATTCCTAAA 780
 DB 4916 GATGGATATAGTACTTAACAGCAGCTCTTTGAGATAGATTTCAGGATAACAATTCCTAAA 4975

QY 781 TCTGATGGAAATTTTATCTTAAGGAAATATAATGATGACACCAAGAAATTCGATATCT 840
 DB 4976 TCTGATGGAAATTTTATCTTAAGGAAATATAATGATGACACCAAGAAATTCGATATCT 5035

QY 841 TTGTCATCTCTTGGCGGGTAAAGTTTAACTCCACAAATGGAACGTGATTAATATT 900
 DB 5036 TTGTCATCTCTTGGCGGGTAAAGTTTAACTCCACAAATGGAACGTGATTAATATT 5095

QY 901 GCTGACGACGCTCTCTGGAACCAAACTGGAATAGAAATTACAGCTGCACCATGCCAGAA 960
 DB 5096 GCTGACGACGCTCTCTGGAACCAAACTGGAATAGAAATTACAGCTGCACCATGCCAGAA 5155

QY 961 ATCAGTGTTCGGGTGTTGTTGTCGCTGACGTTTGCATTTGGAATGGAAGTGGAAAAAT 1020
 DB 5156 ATCAGTGTTCGGGTGTTGTTGTCGCTGACGTTTGCATTTGGAATGGAAGTGGAAAAAT 5215

QY 1021 CCCGAGGCTGACAAATATATGGGTAAATTAATTAATGTTTACTTTCACACCAAGTAGTCAACA 1080
 DB 5216 CCCGAGGCTGACAAATATATGGGTAAATTAATTAATGTTTACTTTCACACCAAGTAGTCAACA 5275

QY 1081 CTCTAG 1086
 DB 5276 CTCTAG 5281

RESULT 4
 AAD55748
 ID AAD55748 standard; DNA; 7239 BP.
 XX
 AC AAD55748;
 XX
 DT 07-AUG-2003 (first entry)
 XX
 DE *Escherichia coli* CS4 operon.
 XX
 KW Coli surface antigen; CS; vaccine; diarrhoea; bacterial infection;
 XX antigen; *csaB*; *csaA*; *csaC*; *csaE*; *ISI*; gene; *ds*.
 XX
 OS *Escherichia coli*.
 XX
 FE Key Location/Qualifiers
 FT misc_feature 283..999
 FT /tag= a
 FT /note= "csaA gene"
 FT misc_feature 1028..1531
 FT /tag= b
 FT /note= "csaB gene"
 FT misc_feature 1589..4192
 FT /tag= c
 FT /note= "csaC gene"
 FT misc_feature 4196..5281
 FT /tag= d

vertebrate against the enterotoxigenic *Escherichia coli*.

Disclosure: Page 35-42 : 58pp; English.

The invention relates to an isolated nucleotide sequence comprising a *csa* operon (encoding proteins required for producing CS4 pili) or its functional fragment. An immunogenic composition comprising a recombinant product of a *csa* operon and a carrier, is useful for generating an immune response in a subject, which involves contacting the subject with the immunogenic composition. The recombinant product of the composition is the CS4 antigen (bacterial pili protein) and is provided in an acellular or cellular preparation. The nucleic acid is useful for producing a polypeptide product from a *csa* operon or functional fragment, which involves providing a nucleic acid, introducing the *csa* operon in an expression vector, such that a recombinant host cell is produced and subjecting the recombinant host cell to conditions such that a protein from the *csa* operon is expressed. The nucleic acid encoding at least an immunogenic portion of the *csa* operon or a polypeptide encoded by the nucleic acid is useful for generating an immune response in a vertebrate against ETEC (enterotoxigenic *Escherichia coli*). The nucleic acid is also useful for identifying polynucleotides encoding other proteins with biological functions similar to that of the *csa* operon and for creating a multivalent Shigella-ETEC immunogenic composition that will protect from diarrhoea caused by either Shigella or CS4 expressing ETEC strains. This sequence represents a *csa* operon related DNA of the invention

Sequence 7239 BP; 2436 A; 1181 C; 1345 G; 2277 T; 0 U; 0 Other; 0 X

Query Match 100.0%; Score 1086; DB 8; Length 7239;

Query Match	100.0%	100%	100%
Best Local Similarity	100.0%	100.0%	100.0%
Pred. No.	1.1e-253	1.1e-253	1.1e-253

Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ACGAATAAGATTTTATTTATTTTACATGTTTTCTCTTCAGTACTTTTACATTTTGGT 60

4196 ATGAAATAGATTTTATTATTTTTACATTGTTTCTCTTCAGTACTTTTACATTGCT 4255

OV 61 GTATCGGCAGATAAAATTCCTGGAGATGAAGCATAACTAATATTTTGGCCCGGTGAC 120

db 4256 GTATCGGCAGATAAAATTCCCGGAGATGAAAAGCATAACTAATAATTTTGGCCCGCGTGAC 4315

121 AGGAACGAATCTTCCCCCAACATAATATATTAATAACCATATTACAGCATACAGTGAA 180

4316 AGGAACGAATCTCCCCCAACATATAATATATTAATTAACCATATTACAGCATACAGTGAA 4375

181 AGTCATACCTCTGATGATAGGATGACCTTTTATGTTTGCTCTCAACAATACACTTAAT 240

4375 AGTCGATACCTCTGATTCATAGGATGACCTTTTATAGCTTCTCTCTCATCAATACACCTTAT 4435

241 GGAGCATGTCCAACTAGTGAGAAATCCTAGCAGTTTCATCGGTGAGGCGGTGAAACAAATATA 300

4436 GCGCGATCTCCAAACCACAGTCGAGCAATCCCTAGCAGTTTCATCGGTCAAGGGTCAAAACAAATATA 4495

301 ACATTACAAATTTCGGAAAAACGAAGTTTAATAAAAAAGAGAGCTACAAATTAAAGGCTAT 360

1186 ACCTTTCAGTTCCGCGCCTATACGGTAAATTAAGGCTAT 4555

Q: 361 A A C A T T A T T C T T C A A A A G T G T T A C T G C C C A T C C G C T A C A C T T A C T C A G C T C A T 420

4615

480

4575

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

[illegible]

C
C
E
F
C
C
K
E
K
E
E
C
T
E
E
K
K
S
C
E
K
K
U
E
A
O
Z
A
O
O
E
U
C
O
O
A
T
T
I
O
N

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PT New bacterial cell expressing three or more coli surface antigens, useful
PT for manufacturing a medicament, i.e. a vaccine, for vaccination against
PT diarrhea.
XX
XX
PS Disclosure; Page 95-97; 58pp; English.
XX The invention relates to a bacterial cell expressing three or more coli
CC surface (CS) antigens. The bacterial cell is useful for manufacturing a
CC medicament, i.e. a vaccine, for vaccination against diarrhoea. The
CC vaccine is also useful for targeting bacterial infection. The present
CC sequence is *Escherichia coli* CS1 operon fragment containing *cooC* and *cooD*
CC gene. This sequence is used in the invention
XX
SQ Sequence 5336 BP; 1553 A; 1021 C; 1206 G; 1556 T; 0 U; 0 Other;
Query Match 28.6%; Score 310.8; DB 7; Length 5336;
Best Local Similarity 58.8%; Pred. No. 2.3e-65;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;
7 AAGATTTATTTATTTTACATGTTTTCTCTTCAGTACTTTTTTACATTTGCTGTATCG 66
2703 AAAAAAGATATTTATTTTGTCTATCATATTTTCTGCGTGTGTCAGTCCCGGGCGATAC 2762
67 GCAGATTAATTTCCGAGATGAAGCATATACTAATATTTTGGCCGCGTGACAGGAC 126
2763 CCGGAATACAGTAGTAACTGACGAAGAGTTTTCAGAGCCCTCTGCTGATAGAGC 2822
127 GAATCTTCCCCCAACATATATTAATTAACCATPATTACAGCATPACAGTGAAGTCA 186
2823 GTACAATACCAATATATAACATCTTTACGAATCAATGGCTGGATATAGTTGAGTCAT 2882
187 ACTCTGTATGAAGAGTACTTTTTTATGTTTGTCTTTCACATACACTTTATGGAGCA 246
2883 AGCTTATATGACAGGATGTTTTTTTATGATACCTCTCGAATCCGGTTATGCTGT 2942
247 TCTCAACAGTGAAGATCCTAGCACTTCAICGGTCAGCGGTCAAAACAATAATATTA 306
2943 TGCCCAACCATTTGGAACATCTGGAGTTCAAT-----ACGGTACTACACCATTA 2996
307 CAATTTACGGAATAAAGAGTTTAAATAAAGAGAGCTACAAATTAAGGCTATAAACA 366
2997 CAGTTTACAGAAAAGAGAGTCTGATAAAGAGAAATTAATCTTGCAGGTAAATAGAA 3056
367 TTATCTTCAAAGTGTAACTGCCCATCCGGCTTAACACTT-----AACTCAGCTCAT 420
3057 CCAATATGGGAGATCAGAGTTCCGACTTTAGCAATCTAATGGTGTGAATTCGAAGTCT 3116
421 TTATCTGTAAATAAAGCG--CGGCTTCAGGTCCAGTTTATATTTATATATCTCTGCT 477
3117 TGGAGCTGTGGGCTCAGCGAATGCTAACGGAACTCTTAATCTGTATATCTCCCTGCA 3176
478 GCGCAATAAAAAATTTGCTTTTGGTGTATCTGGGATCTACTCTGAAGTTAAGAGTA 537
3177 GGAGAAATCAACHAATTTGCTTTTGAGGAGATATGGAGCAACTCTGATCTTACGCTTA 3236
538 AAAG-----ACGATATAGTGAAGTATGGAACCTTACCTATAATATCATATT 588
3237 TCAAGATATGGGAGTCTAGTAGCCATTTACGGCAATATATACCGTAATATTAACGGTT 3296
589 AAATTAACGTAAAGGAATATTCAGATATGTTTACCTCAGTTCAAAGTGAAGTCCGC 648
3297 GATTTAACTGATAAAGGATTAATATTCAGTATGCTTCCAGGGTTTCACAGCAACCGGTT 3356
649 GTCGATCTTAACTTGGTCCCACTGGTGGGGGCACATATATTGGAAGAAATCTGTGTAT 708
3357 GTAGACCTGAATCTGCGCCTATCGGTAAATTAATAATATAGTGTAGTAATTCACCTGAC 3416
709 ATGTGCTTTTATGATGAATATAGTACTACAGCAGCTCTTTTGAGATATAGATTTACGAT 768
3417 ATCTGTGTTCTATGATGGATATAGTACAAACAGGTATAGCATGCTGATAAAGTTCCAGAT 3476
769 AACAAATCCATAATCTGATGGAAATTTTATCTAAGGAATAATATGATGACACCAAGAA 828

Db 3477 GATAATCTACCAATTCATCTGAATATAATCTTTATAAGATA---GGGGGCACTGAAAAA 3533
QY 829 ATTGATATATCTTCTGCTCACTTCTCTTGGGGGTAAAGTTTAACTCCACACAAATGCGACG 888
Db 3534 TTACCATATGCTGTTCATCTGCTTATGGGAGAAAAAATAATTTTATCCAGTGAATGCTCAA 3593
QY 889 TCATTAAATATTTGCTGACGCGACCTTCTCTGGAAAAACAACCTGGAATAGAAATTACAGCTGTC 948
Db 3594 TCATTACTATCAATGACAGTAGTGTACTCGAACAACAACCTGGAATCGAGTAAACCGCAGTT 3653
QY 949 ACCATGCCAGAAATCAGTGTTCGGGTGTTGTTGCGCTGACGTTTGCATTTGCAATTGATGCA 1008
Db 3654 GCTATGCCGGAAGTTAATGTTCCAGTATTATGCTGCCAGCAAGATTGCTATTAAATGCT 3713
QY 1003 AAAGTGGAAATCCCGAGGCTGGCAATATATATGGTGTATATTAATGTTACTTTCACACCA 1068
Db 3714 GATGTAATATGCTCCGATGTCAGGACAGATATTTCAGGACAGATATATATACATTTACACCC 3773
QY 1063 AGTAGTCAAA 1078
Db 3774 AGTGTGAAA 3783
RESULT 7
AAX84848
ID AAX84843 standard; DNA; 5798 BP.
XX
AC AAX84843;
XX
DT 22-SEP-1999 (first entry)
XX
DE CS2 gene cluster.
XX
KW CS2 gene cluster; *CotA*; *CotB*; *CotC*; *CotD*; pilin protein; immunogen;
KW enterotoxigenic *E. coli*; human upper intestine; diarrhoeal disease;
KW enteric infection; therapy; ds.
XX
OS *Escherichia coli*.
XX
PN US932715-A.
XX
PD 03-AUG-1999.
XX
PF 07-JUN-1995; 95US-00483101.
XX
PR 07-JUN-1995; 95US-00483101.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Caron J, Froehlich B, Scott JR;
XX
XX WPI; 1999-443623/37.
DR P-P8DB; AAY22323, AAY22324, AAY22325, AAY22326.
XX
XX Isolated nucleic acids encoding *Escherichia coli* CS2 pilin proteins
PT useful for vaccinating against diarrheal diseases caused by *Escherichia*
PT coli.
XX
XX Claim 2; Col 27-38; 35pp; English.
XX
XX This sequence represents the *Escherichia coli* CS2 gene cluster, and
CC encodes the CS2 pilin proteins (*CotA*, *CotB*, *CotC*, and *CotD*) of the
CC invention. CS2 pilins are long proteinaceous molecules thought to mediate
CC attachment of enterotoxigenic *E. coli* (ETEC) to and/or promote
CC colonisation of the human upper intestine. The CS2 gene cluster may be
CC used to produce immunogens for vaccinating patients against diarrhoeal
CC diseases caused by ETEC bacteria. This type of enteric infection is a
CC major cause of death among infants in developing countries and in
CC immunocompromised (e.g. Acquired Immune Deficiency Syndrome (AIDS)) or
CC elderly adults. The vaccine comprises more than 1 antigenic determinant
CC (epitopes) from more than 1 pilus type to be effective against more than
CC 1 type of ETEC infection
XX

```
SQ Sequence 5798 BP; 1904 A; 874 C; 1221 G; 1799 T; 0 U; 0 Other;

Query Match      27.5%; Score 298.8; DB 2; Length 5798;
Best Local Similarity 58.5%; Pred. No. 1.9e-62;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

QY 127 GAATCTTCCCCAAACATAATATATTAATAACCAATATACAGATGAAAGTCAAT 186
D5 4583 GCATCGTATCCTGCTCATTTATATTTATGTAACATGTTGCTGGTTACAATAAAGATCAC 4642

QY 187 ACTCTGTATGATAGGATGACTTTTATGTTTGTCTCTTCACAAATACACTTAATGAGCA 246
D5 4643 TCTCTTTTTCACAGGATGACGTTTATGATGATCATCAACAGATGATGATGAGCA 4702

QY 247 TGTCCAAACAGTCAGAACTCTAGCAGTTCATCGCTCAGCGGTGAAACAAATATAACATTA 306
D5 4703 TGTCCGACAGGAGAAACTTCCA-----AATCTCTCAAGGGAGACTATATTAAGCTA 4756

QY 307 CAATTTACGGAAAAAGAGATTTTAATAAAGAGAGCTACAAATTAAGGCTATAAACAA 366
D5 4757 ATATTTTACTGAAAAAGAAAGTCTGGCCAGAAAAACATTAACCTTAAAGGATATAAGAGA 4816

QY 367 TTATCTTCAAAAGTGTAACTGC-----CCATCCGGCTACACCTTAACCTCAGCT 417
D5 4817 TTTTATATGAATCAGATAGATGCAATTCATATGTCGATAAATGAATCTCAATTTCTCAT 4876

QY 418 CATTTTAACCTGAATAAAAACGGGCTTCAGGTGCAAGTTTATATTTATATATTTCTGCT 477
D5 4877 ACTGTTAAATGTGTAGTTCAITTCACAGAGGAGTAGATTTCACTTTATATATCCCACAA 4936

QY 478 GCGCAACTAAAAAATTTGCTTTTGTGGTATCTGGGATGCTACTCTGGAAGTTAAGATTA 537
D5 4937 GGTGAATTCATGGCTTCTAATCTGAGGTATATGGGAGGCAACACTAGAGTTACGAGTC 4996

QY 538 AAAGAGCATATAGTGAGACTATGAGACTTACACTATAATATCTACTATTAATTAACCT 597
D5 4997 AAAAGGCATTTACGACATATAATCATGTGTACTTACAAAGTTAATATCACAGTTGATTTGACA 5056

QY 598 GATAAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCAGCTCCGCTGATCTTT 657
D5 5057 GACAAAGGAATATTCAGTCTGGACACCAAGTTTCATAGCGATCTTGAATTTGATCTG 5116

QY 658 AACTTCGCTCCAACTGGTGGGGGACATATATTTGGAGAAATTTCTGTGTATATGTTGTTT 717
D5 5117 AATTTAGCTCCTCAAGGTAATGTTAATATTTCTGTTAGTAACTGCTTGGATGTGCTC 5176

QY 718 TATGATGATATAGTACTACACGAGCTCTTTGAGATATAGATTCAGGATACCAATCCT 777
D5 5177 TATGATGGCTATATAGTACATAGTCAAGTATAGAAATGAGGTTTCAGGATGACTACAA 5236

QY 778 AAATCTGATGGGAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTTGCATAT 837
D5 5237 ACAGGAATATGAATATATCTTATATAAAGTGGAGGCAATTAATAAATTTGCCATAT 5296

QY 838 ACTTTGTCACTTCTTTGGCGGTTAAAGTTTAACTCCACAAATGGAAGCTCATTAAT 897
D5 5297 AAATCTTCTCTTCTTTAGGAGGACGAGATTTTATCCAAATATGAGAGGCTTTTACT 5356

QY 898 ATTCTCAGCAGGCTCTCTCGAAACAACTGGAATAGATTAACAGTGTACATGCCA 957
D5 5357 ATTAATGATATCTCTGTCATTTTATATAACCTGGAATCGTATTAAGTCTGTATCCTTACCA 5416

QY 958 GAAATCAGTGTCCGGTGTGTTGTTGGCTGAGGTTTGAATTTGGAATGCAAAAGTGGAA 1017
D5 5417 CAGATTAGTATTCAGTACTATGCTGCGCCAGCAACTTGACATTTATGTCAGAGCTAAT 5476

QY 1018 AATCCGAGGCTGGACAAATATATCGGTAATATTAATGTTACTTTTCAACCAAGTAGTCAA 1077
D5 5477 AATCCAGAGCGGCTGATATTCAGGATATTAACGTAACATTTACTCTCTAGTAGTTCA 5536

QY 1078 ACATCTTA 1085
D5 5537 AGTCTGTA 5544
```

RESULT 8

```
AD55745
XX AAD55745 standard; DNA; 5798 BP.
AC AAD55745;
XX 07-AUG-2003 (first entry)
DT Escherichia coli CS2 operon.
XX Coli surface antigen; CS; vaccine; diarrhoea; bacterial infection;
KW antigen; cotB; cotA; cotC; cotD; gene; ds.
XX Escherichia coli.
XX Key Location/Qualifiers
FT misc_feature 499..1215
FT /tag= a
FT /note= "cotB gene"
FT misc_feature 1255..1767
FT /tag= b
FT /note= "cotA gene"
FT misc_feature 1836..4436
FT /tag= c
FT /note= "cotC gene"
FT misc_feature 4451..5545
FT /tag= d
FT /note= "cotD gene"
XX WO2003022306-A2.
XX 20-MAR-2003.
XX 11-SEP-2002; 2002NO-GB004123.
XX 11-SEP-2001; 2001GB-00021998.
XX (ACAM-) ACAMBI5 RES LTD.
XX Turner AX, Greenwood J, Stephens JC, Beavis JC, Darsley MJ;
XX WPI; 2003-301009/29.
XX New bacterial cell expressing three or more coli surface antigens, useful
XX for manufacturing a medicament, i.e. a vaccine, for vaccination against
XX diarrhea.
XX Disclosure; Page 99-101; 58pp; English.
XX The invention relates to a bacterial cell expressing three or more coli
XX surface (CS) antigens. The bacterial cell is useful for manufacturing a
XX medicament, i.e. a vaccine, for vaccination against diarrhoea. The
XX vaccine is also useful for targeting bacterial infection. The present
XX sequence is Escherichia coli CS2 operon containing cotB, cotA, cotC and
XX cotD gene. This sequence is used in the invention
XX
SQ Sequence 5798 BP; 1904 A; 874 C; 1221 G; 1799 T; 0 U; 0 Other;

Query Match      27.5%; Score 298.8; DB 7; Length 5798;
Best Local Similarity 58.5%; Pred. No. 1.9e-62;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

QY 127 GAATCTTCCCCAAACATAATATATTAATAACCAATATACAGATGAAAGTCAAT 186
D5 4583 GCATCGTATCCTGCTCATTTATATTTATGTAACATGTTGCTGGTTACAATAAAGATCAC 4642

QY 187 ACTCTGTATGATAGGATGACTTTTATGTTTGTCTCTTCACAAATACACTTAATGAGCA 246
D5 4643 TCTCTTTTTCACAGGATGACGTTTATGATGATCATCAACAGATGATGATGAGCA 4702

QY 247 TGTCCAAACAGTCAGAACTCTAGCAGTTCATCGCTCAGCGGTGAAACAAATATAACATTA 306
D5 4703 TGTCCGACAGGAGAAACTTCCA-----AATCTCTCAAGGGAGACTATATTAAGCTA 4756

QY 307 CAATTTACGGAAAAAGAGATTTTAATAAAGAGAGCTACAAATTAAGGCTATAAACAA 366
D5 4757 ATATTTTACTGAAAAAGAAAGTCTGGCCAGAAAAACATTAACCTTAAAGGATATAAGAGA 4816

QY 367 TTATCTTCAAAAGTGTAACTGC-----CCATCCGGCTACACCTTAACCTCAGCT 417
D5 4817 TTTTATATGAATCAGATAGATGCAATTCATATGTCGATAAATGAATCTCAATTTCTCAT 4876

QY 418 CATTTTAACCTGAATAAAAACGGGCTTCAGGTGCAAGTTTATATTTATATATTTCTGCT 477
D5 4877 ACTGTTAAATGTGTAGTTCAITTCACAGAGGAGTAGATTTCACTTTATATATCCCACAA 4936

QY 478 GCGCAACTAAAAAATTTGCTTTTGTGGTATCTGGGATGCTACTCTGGAAGTTAAGATTA 537
D5 4937 GGTGAATTCATGGCTTCTAATCTGAGGTATATGGGAGGCAACACTAGAGTTACGAGTC 4996

QY 538 AAAGAGCATATAGTGAGACTATGAGACTTACACTATAATATCTACTATTAATTAACCT 597
D5 4997 AAAAGGCATTTACGACATATAATCATGTGTACTTACAAAGTTAATATCACAGTTGATTTGACA 5056

QY 598 GATAAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCAGCTCCGCTGATCTTT 657
D5 5057 GACAAAGGAATATTCAGTCTGGACACCAAGTTTCATAGCGATCTTGAATTTGATCTG 5116

QY 658 AACTTCGCTCCAACTGGTGGGGGACATATATTTGGAGAAATTTCTGTGTATATGTTGTTT 717
D5 5117 AATTTAGCTCCTCAAGGTAATGTTAATATTTCTGTTAGTAACTGCTTGGATGTGCTC 5176

QY 718 TATGATGATATAGTACTACACGAGCTCTTTGAGATATAGATTCAGGATACCAATCCT 777
D5 5177 TATGATGGCTATATAGTACATAGTCAAGTATAGAAATGAGGTTTCAGGATGACTACAA 5236

QY 778 AAATCTGATGGGAAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATTTGCATAT 837
D5 5237 ACAGGAATATGAATATATCTTATATAAAGTGGAGGCAATTAATAAATTTGCCATAT 5296

QY 838 ACTTTGTCACTTCTTTGGCGGTTAAAGTTTAACTCCACAAATGGAAGCTCATTAAT 897
D5 5297 AAATCTTCTCTTCTTTAGGAGGACGAGATTTTATCCAAATATGAGAGGCTTTTACT 5356

QY 898 ATTCTCAGCAGGCTCTCTCGAAACAACTGGAATAGATTAACAGTGTACATGCCA 957
D5 5357 ATTAATGATATCTCTGTCATTTTATATAACCTGGAATCGTATTAAGTCTGTATCCTTACCA 5416

QY 958 GAAATCAGTGTCCGGTGTGTTGTTGGCTGAGGTTTGAATTTGGAATGCAAAAGTGGAA 1017
D5 5417 CAGATTAGTATTCAGTACTATGCTGCGCCAGCAACTTGACATTTATGTCAGAGCTAAT 5476

QY 1018 AATCCGAGGCTGGACAAATATATCGGTAATATTAATGTTACTTTTCAACCAAGTAGTCAA 1077
D5 5477 AATCCAGAGCGGCTGATATTCAGGATATTAACGTAACATTTACTCTCTAGTAGTTCA 5536

QY 1078 ACATCTTA 1085
D5 5537 AGTCTGTA 5544
```



```

Db 4703 TGTCCGACAGGAGAAAACCTCC-----AATCCTCTCAAGGGGAGACTAATATTAGCTA 4756
QY 307 CAATTACGGAAGAAAGATTATATAAAGAGAGCTACAAATTAAGGCTATATAACAA 366
Db 4757 ATATTACTGAAGAAAGAGTCTGCCGAGAAACATTAACCTTAAAGGATATAAGAGA 4816
QY 367 TTAATTGTTCAAAGTGTAACTGC-----CCATCGGCGCTAACACTTAACTCAGCT 417
Db 4817 TTTTATATGAATCAGATAGATGCAATTCATATGTCGATATAAATGAATCTCAATCTCTCA 4876
QY 418 CATTTTAACCTGTAATAAAGACCGGCTTCAGGTGCAAGTTTATATATATATATCTCTGCT 477
Db 4877 ACTGTTAAATGTTAGTTCATTCACAGAGAGTAGATTTCACITATATATATCCACA 4936
QY 478 GCGCAACTAAAAAATTCCTTTGGTGATCTATCGGATGCTACTCTGAGTAAAGTAAAGATTA 537
Db 4937 GGTGAATTTGATGGGCTTCTAAGTCTGAGGTATATGGAGGCAACACTAGATTACGAGTC 4996
QY 538 AAAAGACGATATAGTACAGCTTATGAACTTACACTTATAAATATCACTATTAATTAATTA 597
Db 4997 AAAAGGCATTTACGACTATATATCATGCTACTTACAAAGTTAATATCACTGATTTGACA 5056
QY 598 GATAAGGAAATATTCCAGATATGTTTACCTCAGTTTCAAAAAGTCACGCTCGCTCGATCTT 657
Db 5057 GACAAAGGAATATTCCAGTCTGGACACCAAGTTTCATAGCATCTTGAATTAATGATCTG 5116
QY 658 AACTTTCGCTCCAACTGCTGGGGGACATATATGGAAGAAATTTCTGATATGCTCTT 717
Db 5117 AATTTACGTTCTTGAAGTAACTGTAATATCTGCTGAGTAACTGCTGAGATGCTCTC 5176
QY 718 TATGATGATATAGTACTAACAGCAGCTCTTTGAGATTAAGATTTTTCAGATTAACATCT 777
Db 5177 TATGATGCTATATAGTACATATAGTCAAGATATAGAAATGAGTTTCAGGATGACTCACAA 5236
QY 778 AATCTGATGGGAAATTTTATCTTAAGGAAATATAATGATGACACCAAGAAATTTGCATAT 837
Db 5237 ACAGGAAATATAAGATATAATCTTATAAATCTGAGAGGCCATTAAAAAATTTGCCATAT 5296
QY 838 ACTTTGCTACTTCTCTTGGGGGTAAAGTTTAACTCCAAACAAATGGAAGCTCAATTAAT 897
Db 5297 AAACTTTCTCTCTTTTAGGAGGACGAGAGTTTATCCAAATTAATGGAGAGGCTTTTACT 5356
QY 898 ATTGCTGAGCAGCTTCTCTGAAACAAACTGAAATAGAAATACAGCTGTCCACATGCCA 957
Db 5357 ATTAATGATACTTCGTCATTTGTTTATAAATCGAATCGTATTAAGTCTGTATCCCTTACCA 5416
QY 958 GAAATCAGTGTTCGGGTGTTGTTGGCTCGGACGTTTGCAATGGATGCAAAAGTGGAA 1017
Db 5417 CAGATTAGTATTTCCAGTACTATGCTGGCCAGCAAACTTGACATTTATGTCAGAGCTAAAT 5476
QY 1018 AATCCGAGGCTGGACAAATATATGGGTAAATATTAATGTTACTTTTACACCAAGTATGCA 1077
Db 5477 AATCCAGAGCGGGTGGATATTGAGGATATCTTACGTAATCTTAACTTACTCTCTAGTTTCA 5536
QY 1078 AACTCTTA 1085
Db 5537 AGTCTGTA 5544

```

RESULT 9

ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX AC
XX AC
XX AC

XX 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX plant; bacterial infection; fungal infection; viral infection; rice;
XX gene; ds.

Osryza sativa.
XX WO2003000898-A1.
XX 03-JAN-2003.
XX 22-JUN-2001; 2001WO-IB001105.
XX 22-JUN-2001; 2001WO-IB001105.
XX (SYN) SYNGENTA PARTICIPATIONS AG.
XX Chang H. Chen W. Cooper B. Glazebrook J. Goff SA, Hou Y;
XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
XX Identifying at least one gene involved in plant resistance or response to
XX pathogenic infection for conferring resistance or tolerance to a plant to
XX bacterial, fungal or viral infection by determining or detecting plant
XX gene expression.
XX Claim 27; SEQ ID NO 5263; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
XX involved in plant resistance or response to pathogenic infection. M1
XX comprises identifying a gene whose expression is significantly altered in
XX the incompatible interaction of plant gene expression relative to
XX the incompatible interaction of plant gene expression relative to
XX expression of the gene in an uninfected plant, in a mutant plant that
XX does not express a gene associated with response to pathogenic infection,
XX or in a corresponding incompatible or compatible interaction. (M1) is
XX useful for conferring resistance to resistance or tolerance to a plant to
XX bacterial, fungal or viral infection. The present sequence was used to
XX illustrate the invention.

XX SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.6%; Score 50; DB 7; Length 2000;

Best Local Similarity 9.0%; Pred. No. 0.04;

Matches 61; Conservative 318; Mismatches 298; Indels 3; Gaps 2;

QY 221 CTCTCACAAATACCTTAATGAGCAGCATGTCACCAACAGTCCAGAAATCTAGCAGTTCTCGG 280
Db 1019 CTGWRVYFRSNTWYAMKVKYTKYTAYSWTWKYWAYKRAYAKSRKTKWCCTGGK 960
QY 281 TCAGCGGTGAACAAATATAACATTACATTTACGGAAGAAAGAACTTTAATAAAGAG 340
Db 959 MATYCTCTKWAJGRWRWRWAWCWYCCMKWKKVTSOMWKYRTWCWYTMWGAAYAY 900
QY 341 AGTACAAATTAAGGCTATATAACAATTTATGTTCAAAAGTGTTAACCTGCCATCCGGCC 400
Db 899 YAMRRRTYKWSWRMYWTWKWATWMTCMCAKMYATGATWATWWRVYTWYTYCYAM 840
QY 401 TAACACTTAACCTCAGCTCATTTTAACGTGAATAAAGAACGGC-GCTTCAGGTGCAAGTTTA 459
Db 839 TCAKCKYKMAATKWTWTWACARATSWRWEZAGMEWKYKMKRAYWRWRWRCWAGWAR 780
QY 460 TATTATATATCTCTGCTGCGNACTAAAAATTTCCCTTTTGGTGTATCTGGATGCT 519
Db 779 WMKSRVYRWKKYATRYKWKWAMTWWSWRWKSYSRMSGMRMWSAWRYCSRWCKAK 720
QY 520 ACTCTGAAGTTAAGATAAAGACGATATAGTACACCTTATGAGACCTTATGACATATAAAT 579
Db 719 TKVASSARWTKAKRSYRYYRRMYWKRGNTYRYWRSCRWTRARYSK--RRKWAGASM 662
QY 580 ATCACTATTATAATTAACCTGATAGGGAATATTCAGATATGTTTACTCAGTTCAAAAGT 639
Db 661 KSCWMYWRGARSWYSKYSKSAKCKCKTRYMTSSYMSVTGMYKYSYKMSWTSKMSYMG 602
QY 640 GACGCTCGGTGATCTTAACTTCGCTCCAACTGGTGGGGGCACATATATTCGAAGAAAT 699
Db 601 KMTCTMYTSMKGTSTRSKGRWSGMSRMYRWWKMKRKYMKYMKWCKTWRRCMYRWG 542

CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours
CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-AB:92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 10279 BP; 3242 A; 65 C; 1845 G; 5127 T; 0 U; 0 Other;
Query Match 4.5%; Score 49.4; DB 6; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.089;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
QY 140 AACATTAATATTAATAACCATATTTACAGCATACAGTGAAAGTCATCTCTGTATGATA 199
DB 7605 ATCATATATAAATAATTTAAATAAATAAACAATAAATAATTAATCCAAAAACA 7546
QY 200 GGATGACATTTTATGTGTTCTTCACATACACTTAATGAGCATGTCACACAGTG 259
DB 7545 ATAAACCATTAATTAACATTAATTTTCCAAATATCCAAAAAATCTTACTCTCAATA 7486
QY 250 AGATCCTAGCGTTTCATCGGTCAGCGGTGAAACAAATATACATTAATTTACGGAAA 319
DB 7485 AAAAAAATAAATACTTAATTTCAACACGAAATAAATAAATAAATAAATAAATAA 7426
QY 320 AAAGAAGTTTATAAATAAAGAGAGCTACAAATTAAGGCTATAAACAATTTATGTTCAAAA 379
DB 7425 ACATATAAATACTAATATATATACAAATCCAACTTTTATCAACAAACAATTAATAA 7366
QY 380 GTGTTAACTGCCATCCCGGCTTAACACTTAATCTAGCTCATTTTACTGTATATAAAC 438
DB 7365 ACTATATCTTCATAAACACCCCAATATTTTATATATCTCATTTTAAATTCATTAAATC 7307
RESULT 12
AAD22328/c
ID AAD22328 standard; DNA; 10279 BP.
XX
AC AAD22328;
XX
XX 12-FEB-2002 (first entry)
XX Chemically treated human genomic DNA #18 associated with DNA adducts.
XX DNA adduct; peptide nucleic acid; PNA; cytosine methylation;
XX Gene therapy; tumour; cancer; human; ds.
XX Homo sapiens.
XX WO200177378-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP004015.
XX
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2003DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-010923/01.
XX Novel nucleic acid comprising sequence of a segment of chemically
XX pre-treated DNA of genes associated with DNA adduct, useful for diagnosis
XX and therapy of solid tumors and cancer.
XX

PS Claim 1; Page 63-68; 97pp; English.
XX The invention relates to chemically modified DNA of genes associated with
CC DNA adducts, oligonucleotides and/or peptide nucleic acid (PNA) oligomers
CC for detecting cytosine methylations as well as method for ascertaining
CC genetic and/or epigenetic parameters of genes associated with DNA
CC adducts. Oligomers of the invention coupled to a solid phase is useful
CC for manufacturing an arrangement of different oligomers (array) fixed to
CC a carrier material for analysing diseases associated with the methylation
CC state of the CpG dinucleotides of chemically pretreated DNA of genes
CC associated with DNA adduct. They are also useful for ascertaining genetic
CC and/or epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases by analysing cytosine
CC methylations. Sequences of the invention are useful for diagnosis and
CC therapy of solid tumours and cancers. They are also useful in gene
CC therapy. The present sequence is chemically pretreated human genomic DNA
CC associated DNA adducts
XX
SQ Sequence 10279 BP; 3242 A; 65 C; 1845 G; 5127 T; 0 U; 0 Other;
Query Match 4.5%; Score 49.4; DB 6; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.089;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
QY 140 AACATTAATATTAATAACCATATTTACAGCATACAGTGAAAGTCATCTCTGTATGATA 199
DB 7605 ATCATATATAAATAATTTAAATAAATAAACAATAAATAATTAATCCAAAAACA 7546
QY 200 GGATGACATTTTATGTGTTCTTCACATACACTTAATGAGCATGTCACACAGTG 259
DB 7545 ATAAACCATTAATTAACATTAATTTTCCAAATATCCAAAAAATCTTACTCTCAATA 7486
QY 260 AGAATCCTAGCGTTTCATCGGTCAGCGGTGAAACAAATATACATTAATTTACGGAAA 319
DB 7485 AAAAAAATAAATACTTAATTTCAACACGAAATAAATAAATAAATAAATAAATAA 7426
QY 320 AAAGAAGTTTATAAATAAAGAGAGCTACAAATTAAGGCTATAAACAATTTATGTTCAAAA 379
DB 7425 ACATATAAATACTAATATATATACAAATCCAACTTTTATCAACAAACAATTAATAA 7366
QY 380 GTGTTAACTGCCATCCCGGCTTAACACTTAATCTAGCTCATTTTACTGTATATAAAC 438
DB 7365 ACTATATCTTCATAAACACCCCAATATTTTATATATCTCATTTTAAATTCATTAAATC 7307
RESULT 13
ABK28407/c
ID ABK28407 standard; DNA; 9504 BP.
XX
AC ABK28407;
XX
XX 23-APR-2002 (first entry)
XX
XX DNA transcription associated genomic DNA #141.
XX
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
XX PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
XX single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
XX viral infection; Sezary syndrome; haematological disorder; tuberculosis;
XX immunological disorder; Werner syndrome; developmental disorder;
XX psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
XX neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
XX myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
XX angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
XX polyglutamine disorder; solid tumour.
XX
XX Unidentified.
XX
XX WO200192565-A2.
XX
XX 06-DEC-2001.
XX
XX 06-APR-2001; 2001WO-EP003973.
XX

XX 06-APR-2000; 2000DE-01019058.
 PR 07-APR-2000; 2000DE-01019173.
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX (EPIC-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-090046/12.
 XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological disorders,
 PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
 PT cancer.
 XX Claim 1; SEQ ID NO 281; 32pp; English.
 XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC disease, HDR syndrome arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28121-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification but
 CC was obtained in electronic format directly from the European Patent
 CC Office
 XX Sequence 9504 BP; 2528 A; 138 C; 183 G; 4985 T; 0 U; 0 Other;
 SQ Query Match 4.4%; Score 48.2; DB 6; Length 9504;
 Best Local Similarity 46.8%; Pred. No. 0.17;
 Matches 152; Conservative 0; Mismatches 173; Indels 0; Gaps 0;
 QY 317 AAAAAGAAGTTTAAATAAAGAGAGAGCTACAACTAAAGCGCTATAACAATTTATTGTCA 376
 DB 9072 AACATAAAATTTCTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9013
 QY 377 AAAGTGTAACTGCCATCCGCTCAACCTTAACCTCACTCACTCACTCACTCACTCACTCACT 436
 DB 9012 AANTAGACACAAATAAACAATCTAAACAATTAATTAATAATAAATAAATAAATAAATAA 8953
 QY 437 ACGCGGCTTCAGGTGCAAGTTTATATTTATATATTTCTGTGCGCACTAAAAAATTTGC 496
 DB 8952 AAACAAAACGGTTTAACTAAATTTATATAAATAAATAAATAAATAAATAAATAAATAA 8893
 QY 497 CTTTTGTGTGTAICTGGATGCTACTCTGAAGTAAAGTAAGTAAGTAAGTAAGTAAGTAAG 556
 DB 8892 TTATAAAAAACATCACTATAAATAAATTCATAAATTAATAAATTTATCTTTTAAACACTAAA 8833
 QY 557 CCTATGGAACCTACACTATATAAATCTACTATTAATTAATTAATTAATTAATTAATTAATTA 616
 DB 8832 AATCACTAAAAAATAAATCAATCTTCTTAACAAATAAATCACTTAACACTATAAATAA 8773
 QY 617 TATGGTTTACCTCAGTTCAAAGTGA 641
 DB 8772 AATATATCTCTCATTTCTTAATAATA 8748

RESULT 14

AAA70139
 ID AAA70139 standard; DNA; 5139 BP.
 AC AAA70139;
 XX 07-NOV-2000 (first entry)
 DT Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:272.
 DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
 XX Plasmodium falciparum.
 OS WO2000025728-A2.
 PN 11-MAY-2000.
 PD 05-NOV-1999; 99WO-US026796.
 XX 05-NOV-1998; 98US-0107131P.
 PR (HOFF/) HOFFMAN S.
 XX (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX Hoffman S, Carucci D, Gardner M, Venter JC;
 XX WPI; 2000-365347/31.
 XX Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.
 XX Disclosure; Page 482-484; 577pp; English.
 PS The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX Sequence 5139 BP; 2440 A; 434 C; 712 G; 1553 T; 0 U; 0 Other;
 SQ Query Match 4.3%; Score 47.2; DB 3; Length 5139;
 Best Local Similarity 44.5%; Pred. No. 0.25;
 Matches 187; Conservative 0; Mismatches 233; Indels 0; Gaps 0;
 QY 193 TATGATAGATGACTTTTTTATGTTTCTCTTCACAATACACTTAATGAGCATGTCCA 252
 DB 997 TATTATAATAATATATTTTAAATCCATTAATGATAATTCATTAATGTTGTTATTACA 1056
 QY 253 ACCAGTGAAGAACCTAGCAGTTTCATCGGTACGGTGAACAACAATATACATTACATTT 312
 DB 1057 ACTTGTGATAAAGAAAGATTATCGAAGCAGCACAAGTGATATGATAAACAATTTTT 1116

QY 313 ACGGAAAAAGAGTATTATAAAGAGAGCTACAAATTAAGGCTATAACCAATTATTG 372
 Db 1117 GAAAGATGATGATGAGAGAAAAATAATATATTAAGAAATGAAAAATTTTATGATGTTATT 1176
 QY 373 TTCAAAAGTGTAACTGCCCATCGGCCTPAACACTTAACTTCAGCTCATTTTAACTGTAAT 432
 Db 1177 AAAAAAATAATCAATGCGCAAGAGAAAAATACTTAAATCAATAGTACAAATTAATAG 1236
 QY 433 AAAAAAGCGGCTTCAGGTGCAAGTTTATATTTATATATTCCTGCTGCGAAGCTAAAAAT 492
 Db 1237 AAAAAAGGTTTC-TTTTATATATTTCTAAGGATGAGGACTTTTAAAGAAAAAGAAAAATAGT 1296
 QY 493 TTGCGCTTTTGGTGTCTCTGGGATGCTACTCTCAAGTTAAGAGTAAAAAGACGATATAGT 552
 Db 1297 TATAATATGGTGTAAAGAGGGAACCCAGGAGATATAATGTAATAAAAAATAATATG 1356
 QY 553 GAGACTATGGAACTTACACTATAATATCACTATTAATTAATTAAGTAAAGGGAATATT 612
 Db 1357 AAAAGCAATAATAAATATTGATAATAATGATAATAATAATAATAATAATAATAATAATAA 1416

RESULT 15

ABLI0866/c

ID ABLI0866 standard; cDNA; 4176 BP.

AC ABLI0866;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27080.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66763.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Claim 1; SEQ ID NO 27080; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 4176 BP; 1186 A; 862 C; 837 G; 1291 T; 0 U; 0 Other;

Query Match

4.2%; Score 45.2; DB 4; Length 4176;

Best Local Similarity 56.6%; Pred. No. 0.73;
 Matches 103; Conservative 0; Mismatches 78; Indels 1; Gaps 1;
 QY 689 TTGGAAGAAATTTCTGTTGACATGTGCTTTTATCATGGAATATAGTACTAACAGAGCTCTT 748
 Db 2504 TCGAAAGCAATATATTTTAAATATGCTTCTGGTATGGTTAAACATCAAAAAGAGCATTTT 2445
 QY 749 TGGAGATAGATTTCAGGATACCAATCCCTAAATCTGATGGGAAATTTTATCTAAGGAAA 808
 Db 2444 AGCAATACGATTCGATGCTAATAACCCGTCCGTTCGAAAAATATTACCATATGAAC 2385
 QY 809 TAAATGATCACACCAAGAAATTTGCATATACT-TTGTCACTTCTCTTGGCGGTAAAGT 867
 Db 2384 TAAATTGAACACTAAGAACTTTAATATTCCTCTAATTAATCTCTTCGCCAGAAAAAT 2325
 QY 868 TT 869
 Db 2324 TT 2323

Search completed: May 8, 2004, 09:40:48
 Job time : 495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 09:32:30 ; Search time 3220 Seconds
(without alignments)
10071.530 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: atgaataagattttattat.....caagtagtcaaacactcttag 1086

Scoring table: IDENTITY NJC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_fod:*
26: em_gss_png:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	59.4	5.5	1101	CNS0100X	AL098379 Drosophila
	55	5.1	1056	13	EX415058
	52	4.8	1200	13	EX415878
c 4	51.6	4.8	1101	29	CNS0039G

RESULT 1
CNS0100X/c
LCUS
DEFINITION
Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION
AL098379
VERSION
AL098379.1 GI:5605990
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPIH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payar. It has been constructed in the vector

ALIGNMENTS

5	51.2	4.7	1067	14	CD386564
c 6	51	4.7	1201	29	CNS00FVZ
c 7	51	4.7	1225	29	CNS0161D
c 8	50.8	4.7	1124	13	EX436282
c 9	50	4.6	1179	13	EX380916
c 10	49	4.5	1043	29	CNS0145P
c 11	49	4.5	1061	13	EX437039
c 12	48.6	4.5	1201	13	EX446296
c 13	48.4	4.5	1201	13	EX355654
c 14	48	4.4	1101	29	CNS0039G
c 15	47.6	4.4	931	14	CD389151
c 16	47.6	4.4	999	13	EX380865
c 17	47.6	4.4	1101	29	CNS0145U
c 18	47.6	4.4	1148	13	EX360752
c 19	47.4	4.4	925	28	BH152177
c 20	47.4	4.4	1101	29	CNS000D1
c 21	47.2	4.3	1101	29	CNS014XJ
c 22	47.2	4.3	1204	29	CNS016E2
c 23	47	4.3	838	29	CNS0096O
c 24	47	4.3	1101	29	CNS00FMC
c 25	47	4.3	1194	13	EX462185
c 26	46.8	4.3	536	13	BQ400483
c 27	46.8	4.3	1101	29	CNS016HF
c 28	46.6	4.3	438	10	BE224677
c 29	46.6	4.3	460	12	BI772835
c 30	46.6	4.3	1201	9	AL532454
c 31	46.6	4.3	1201	13	EX343343
c 32	46.4	4.3	812	28	AZ529785
c 33	46.4	4.3	911	28	AZ687106
c 34	46.4	4.3	1092	29	CNS020K7
c 35	46.4	4.3	1101	29	CNS000B8
c 36	46.4	4.3	1101	29	CNS00EVL
c 37	46	4.2	427	13	EX403499
c 38	46	4.2	1026	13	EX415339
c 39	45.8	4.2	1201	9	AL536104
c 40	45.6	4.2	852	29	CNS009AX
c 41	45.6	4.2	928	29	CNS00DKY
c 42	45.2	4.2	903	29	CNS0173V
c 43	45.2	4.2	994	13	EX414650
c 44	45.2	4.2	1080	29	CNS005PF
c 45	45	4.1	764	14	CF289381

CNS010CX 1101 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL098379

AL098379.1 GI:5605990

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPIH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payar. It has been constructed in the vector

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Séquençage
BP 191 91005 EVRY cedex - France
Email: seqr5@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies
Invitrogen. Contact : Peng Liang Email : fli.invitrogen@fulllength.invitrogen.com/
<http://fulllength.invitrogen.com/> Invitrogen
Paradise Avenue Genoscope sequence ID : CS0CA
Location/Qualifiers

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

1. 1200
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP008Y104"
/tissue_type="THYMUS"
/clone_id="Homo sapiens THYMUS"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched, into
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

```

ORIGIN

Query Match	4.8%;	Score 52;	DB 13;	Length 1200;
Best Local Similarity	35.9%;	Fred. No. 6.1;	Mismatches 75;	Indels 3; Gaps 2;
Matches 229;	Conservative			

QY	4	AATAAGATTATTTATTTTACATCTGTTTTCTCTTCAGTACTTTTACATTTGCTGTGA	63
Db	401	ADDCITATWTWHCCATSTTSCTTTTAAAAATTTTBBSSTWATATTTAATTTTATTA	460
QY	64	TGGCGAGATAAAATTCGGGAGATGAAGCATCACTAATATTTTGCGCGCGTGA CAGG	123
Db	461	TAATTTATTTWATTTTAAATATATTTTTHAAAATTTTTPAATTTTAAAAATTTWATANWWT	520
QY	124	AACGAATCTCCCCCAAACAATAATATATAATAACCATAATCAGGCATACAGTGAAGT	183
Db	521	TWAWATTTTTPAATTWAAATTAATWATTTTTTTTTTAAATATTTTAAAAAAWAATWATTT	580
QY	184	CATACCTGTATGATAGGATGACTTTTTTATGTTGTTCTCACAACTACACTTAATGGA	243
Db	581	TWTTTTTTWAAAATTTTATTTTWTWTWTATTTTTTATTTTATTTATTTATTTAATAA	640
QY	244	GCAATCTCAACACAGTGA - GAATCCTAGCAGTTCATCGGTCAGCGCTCAACAAATATAA	301
Db	641	ATWAAATATAATTTAAWTWRANWAAWTWATATTTTTTTAATAGATTAATATATATAT	700
QY	302	CATTACAAATTCGGGAAAAAGAAGTTTAAATAAAAGAGCGCTACAAATTAAGGCTATA	361
Db	701	AAATTTATTTATTTWAAATWWWATTTTWTAYAGAAATTAATTAATCATATATATAIA	760
QY	362	AACAAATATGTTCAAAAGTGTAACTGCCCATCTCGGCTAACACATTAACCTCAGTCAAT	421
Db	761	TAWDCATRAAAATATWATWTWTATTAATAATTTTAWAAATWCASAAATWAAAAAAATTTAA	820
QY	422	TTAACTGTATAAAAAACGGCGTTCAGGTGCGAAGTTTATTTATTTATATATCTCTGCTGGCG	481
Db	821	AAAATTTTAAATTTWAAAAAFWAATTTTAAAAAAAATTTTATTTTAAATTTWATWATTTA	880
QY	482	AACATAAAAAATTTGCCTTTTTGGTGGTATCTCGGATGCTACTCTCGAAGTTAAGAGTAAAAA	541
Db	881	TAATWATAAAATTAGAAAAATTTTTTTTWTATTTATTTTATTTTAAATTTWAAATTTT	940
QY	542	GAGCATATAGTGAGACCTACGGAACTTACACTATAATATCACTATTAATTAACCTGATA	601
Db	941	-ATATAATAWAAWAAAAAAATTTTWWAAWWTATAAAATANWTATWAAATWAAAAATWW	999
QY	602	AGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGT	639
Db	1000	ATWAAWTTWAAAAATWTATTTTWTWTWTTTAAAAAAAATTT	1037

RESULT 4	CNS0039G/c	LOCUS	DEFINITION
----------	------------	-------	------------

CNS0039C 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TERT end of BAC #
BAC08K10 of RPCL-98 library from Drosophila melanogaster (fruit
fly) genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL063921
AL063921.1 GI:4941778
GSS.
Drosophila melanogaster (fruit fly)
GSS.
Drosophila melanogaster
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The BDGP Drosophila](http://www.fruitfly.org/The%20BDGP%20Drosophila) melanogaster BAC library was prepared by Kazutcyo Oseegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and RST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://www.fruitfly.org/bdgp/bdgp.html>.

FEATURES
SOURCE

```

location="Quarantini"
i="1101"
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACF08K10"
/clone_lib="RPCI-98"
/contig="end : 733"

```

ORIGIN

Query Match	4.8%	Score 51.6;	DB 29;	Length 1101;
Best Local Similarity	16.1%;	Pred. No. 7.5;	Indels 0;	Gaps 0;
Matches 115;	Conservative 280;	Mismatches 319;	Indels 0;	Gaps 0;
139	AAACATAATATTAATAAACCATATATTACAGCATACAGTGAAGCTATACACTCTCTGTATCAT	198		
1098	RRAGDDTWBDDTRKDDWDKWTWTKWDRADRRRWAGDADRWMDGAGTWTATWWWW	1039		
199	AGATGACCTTTTATGTTTGTCTTCTCACATACACTTAATGGAGCATGTCACCAACAGT	258		
1038	WWATWDTWDKWWMTAAKTDATWRTAWRADWAGRDGAGGRDRDAATDADGAGR	979		
259	GAGAACTCTAGCAGTTTCATCGGTGACGGGTGAACCAATATAACATTAACAATTATTCAGCAA	318		
978	RDGGRKKRKRRKGGDDDKKGGKKKAAGKWAATKWWDJWDWDKWDGAKCRKADD	919		
319	AAAAGAAGTTTAAATAAAAGAGAGCTACAAATTAAAGGCTATAAACAATTATTGTTCAA	378		
918	DDGAGKDDDKGKADADDTDGCTKDDDDKDKDDDDKAGTGMGDTATWAAATDWWWWGWA	859		
379	AGTGTTAACTGCCCATCCGGCCTTACACTTAACCTACAGTCTATTTTAACTGTAAATAAAAAC	438		
858	DADKWTWDAAADDWADDWDWAWKWDWDAAWGAATADRRDWGDBAGKGGAKRRDRK	799		
439	CGCGCTTCAGTCCAAAGTTTATATTATATATCTCCCTGCTCGCGCAACTAAAAAATTGCGCT	498		


```

Db 798 RADDKRDADDDDAATWTTTTRDTDDWKKTDTWTRWAADRTWDRDDDDDRDRAG 739
Qy 499 TTTGCTGGTATCGGAGTACTCTGAGTACAGTAAGAAAGACGATATAGTGAGACC 558
Db 738 TAGKWRRTWKRRWKRRDTRWDADADTDDDDRRRRGGGADAGKGTGRKRRDR 679
Qy 559 TATGGAACTACACTATAATATCACTATTAATTAATTAAGGGAATATTCAGATA 618
Db 619 TWDRTDAWADAATWTTTDTDDDRDRKRGARRRRRTTARAANDWWTWKAWDNAKW 619
Qy 619 TGGTACTCAGTCAAAAGTACGCTCGCTCGATCTTAACITTCGCTCCAACTGGTGGG 678
Db 618 DWKTRDRWRWAADTWTDDAKARDWAKARAWASDRDRAABARDBRWTTKGTATTATW 559
Qy 678 GGCACATATATGAAGAAATCTGTCATATGCTTTTATGATGGAATAGTACTAAC 738
Db 558 TTWAAARAANWAAWATTAATWTTTWTWTTTWTWTTTWTWTTTAAWAAWTTAA 499
Qy 739 AGCAGCTCTTTGGGAGATAAGATTTTCAGATTAACATCTGAATCGGGAATTTAT 798
Db 498 WTAANWAAAAAATTTTWTWTTTWTWTTTWTWTTTAAWAAWTTTWTWAAATTTT 439
Qy 799 CTAAGGAATAATATGATGACACCAAGAAATTCATATATATTTGTCACTTCTC 852
Db 438 TWTWTTWAAATTAATTTTWTWTTTWTWTTTWTWTTTAAWAAWTTTAAWAAWTTAA 385

```

```

RESULT 5
CD386564
LOCUS CD386564 1067 bp mRNA linear EST 30-MAY-2003
DEFINITION AGENCOURT_14286445 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CD386564
VERSION CD386564.1 GI:31222277
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1067)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Daniela S. Gerhart, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Dr. Jamie Thompson, University of WI
CDNA Library Preparation: Gina Zastrow-Hayes
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDKM74 row: e column: 02
High quality sequence start: 8
High quality sequence stop: 462.
Location/Qualifiers
1. 1067
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="embryonic trophoblasts, made from WA01 stem
cells"
/lab_host="DE10B Tona"
/clone_lib="NIH_MGC_173"
/note="Vector: pDONR201; Site 1: attP2; Site 2: attP1;
LIBR_PRIMING: oligo dt; METHOD - full-length enriched;
LIBR_PROVIDER - Stratfield"

```

```

FEATURES
source
1. 1067
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="embryonic trophoblasts, made from WA01 stem
cells"
/lab_host="DE10B Tona"
/clone_lib="NIH_MGC_173"
/note="Vector: pDONR201; Site 1: attP2; Site 2: attP1;
LIBR_PRIMING: oligo dt; METHOD - full-length enriched;
LIBR_PROVIDER - Stratfield"

```

ORIGIN

```

Query Match 4.7%; Score 51.2; DB 14; Length 1067;
Best Local Similarity 44.3%; P-adj. No. 9;

```

```

Matches 155; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
Qy 290 AACAATAATACATTTACAATTTACGGAATAAGAGTAAATAAGAGAGCTACAAA 349
Db 351 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 410
Qy 350 TTAAGGCTATAACATTTATTTGTTCAAAAGTGTAACTGCCCATCCGGCTAACCTTA 409
Db 411 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 470
Qy 410 ACTCAGCTCATTTTAACTGTAATAAATAAATCGCGCTTCAGGTGCAAGTTTATATATATA 469
Db 471 ANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 530
Qy 470 TTCTGCTGCGAACTAAAAAATTTTGGCTTTTGGTGTATCTGGGATGCTACTGAGT 529
Db 531 TTTTNTTTTNNAAAAAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 590
Qy 530 TAAGAGTAAAAAGCAGATATAGTGAGACCTATGAGACTTACACTATAAATATCATTA 589
Db 591 TTATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 650
Qy 590 AATTAACTGTAAGGGAATAATTCAGATATGGTTACCTCAGTTCAAAAGT 639
Db 651 TAATNTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATT 700

```

```

RESULT 6
CNS00FVE 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T3T3 end of BAC:
DEFINITION BACR32P18 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL071298
VERSION AL071298.1 GI:4951138
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999); Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aron Mammosser in Piter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR32P18"
/clone_lib="RPCI-98"
/note="end : T3T3"

```

FEATURES

source

ORIGIN

```

Query Match 4.7%; Score 51; DB 29; Length 1101;

```

[illegible]

```

RESULT 8
BX436282/c
LOCUS
DEFINITION Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP0011901
5-PRIME, mRNA sequence.
ACCESSION BX436282
VERSION BX436282.1 GI:3078752-
ES1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1124)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/
Paradey Avenue Genoscope sequence ID : CS0CAP001AB01QPI.
Location/Qualifiers
1..1124
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP001YCC1"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
with a NotI-cligod(NT) primer. Five prime end enriched, int
double-strand cDNA was digested with Not I and cloned, int
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized"
FEATURES
source

```



```

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Feng Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/invitrogen/1690
Paraday Avenue Genoscope sequence ID : XCL0BA900-ZA03FP1.
Location/Qualifiers
1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="XCL0BA900:ZA03"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="Vector: pCMVSPORT6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

```

CRIGIN

Query Match	4.5%;	Score 48.6;	DB 13;	Length 1201;
Best Local Similarity	35.0%;	Pred. No. 25;		
Matches 129;	Conservative 53;	Mismatches 187;	Indels 0;	Gaps 0;
QY	11	TTTATTATTTTACATCTGTTTTCCTCAGTACTTTTACATTTGCTGATCGCAG	70	
Db	152	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	211	
QY	71	ATAAAATCCCGAGATGAAAGCATAACTAATATTTTGGCCCGCTGACAGGAACGAAT	130	
Db	212	WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW	271	
QY	131	CTTCCGCCCAACATATATATTATTAATTAACCATATTACAGCATACAGTGAAGTCATACT	190	
Db	272	ATATATATATATATATATATATATATATATATATATATATATATATATATATAT	331	
QY	191	TGTATGATGAGATGACTTTTTTTATGTGTCTCTCAATATACACTTAATGGACATGTC	250	
Db	332	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	391	
QY	251	CAACCGTGTGAAATCCTTAGCAGTTCATCGGTCAGCGGTGAACAATATATACACTTACAAT	310	
Db	392	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	451	
QY	311	TTACGCAAAAAAACAAGTTTAAATAAAGACAGACTACAAATTAAGGGCTATAACAATAT	370	
Db	452	GRRAAATAAAGGRRAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	511	
QY	371	TGTTCAAAA	379	
Db	512	GRGGRAAAA	520	

RESULT	13
LOCUS	BX355654/c
DEFINITION	BX355654 Homo sapiens PLACENTA COT 25-NORMALIZED Homc sapiens CDNA clone SC0D1C02YE12 3-PRIME, mRNA sequence.
ACCESSION	BX355654
VERSION	BX355654.1 GI:30371992
KEYWORDS	EST.
SOURCE	Homo sapiens
ORGANISM	Homo sapiens (human)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seq-ref@genoscope.cns.fr Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com/> InvitroGen Corporation 1600 Faraday Avenue Gencscope sequence ID : CS0DI002BC06NP1.

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DI002YE12"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT6 vector. Library was normalized."

ORIGIN

Query Match	4.5%; Score 48.4; DB 13; Length 1201;
Best Local Similarity	29.3%; Pred.No. 27;
Matches 71; Conservative 72; Mismatches 99; Indels 0; Gaps 0	
QY	226 CACAATACACTTAAATGGAGCATGTCCAACGAGTCAGAATCCTAGCAGTTTCATCGGTGCAGC 285
DB	1201 HMMVAMMVAHMMMGGMWMTITWAKNMAATTAAWATTTAAMAATTTWAWATTHHHAH 114
QY	256 GGTCGAACAATAATACATTACATTTACGGAAAAAGAGCTTTAATAAAAAGCAGCTA 345
DB	1141 MAHNHTHTTTTTMMMWTTTTTGMTWTWNAWAAAAAAMATTWATAATWAAAAAMMTTMC 108
QY	346 CAATTAAGAGGCTATAACAATATTCTCAAAAGTGTTAACTGCCCATCGGGCTAACA 405
DB	1081 MWTTHTHMNTMHMMMGNGSATTTTWTAAAAAATTTWTAATTTTATTMMWMMTTW 102
QY	406 CTTAACCTACGCTCATTTTAACTGTAAATAAAAACGGGCTTCAGGTGCAGTTTATTTA 465
DB	1021 HMTMTTMAHTMTTTHHAHTTWAATTTWTTTAAAAATTAATTTWTTMA'TWHATTTAAA 962
QY	466 TA 4E7
DB	961 AA 960

RESULT 14	CNS0039G	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	CNS0039G				
DEFINITION	Drosophila melanogaster genome survey sequence TPT3 end of BAC # BACR06K10 of RCI-98 library from drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL063921				
VERSION	AL063921				
KEYWORDS	GI:4941778				
SOURCE	GSS.				
ORGANISM	Drosophila melanogaster (fruit fly)				
	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Ooeegawa and Aaron Mammossier in Piter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial <i>NotI</i> digestion of Drosophila DNA provided by the BDGP from the				

isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

FEATURES	
SOURCE	

```
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACROSK10"
/clone_lib="RPCI-98"
/note="end : TAT3"
```

ORIGIN

	Query Match	4.4%	Score 48	DB 29	Length 1101	
	Best Local Similarity	13.76%	Pred. NO. 33	183	Indels 0	Gaps 0
	Matches 64	Conservative 215	Mismatches			
Qy	5	ATAAGATTATTTATTTTATTTTATACATTGTTTCTCTTCAGTACTATTTTACATTTTGCCTGAT	64			
Db	612	HTYAMWHEWTHWITMAWHEWTVTAAYYYUYICMYYYHYMHHAHAABAAWVTHHTW	671			
Qy	65	CGGCAGATAAATATCCCGGAGATGAAGAGCATACTAATATTTTTTGGCCCGGTCACAGGA	124			
Db	672	THAYHWATYUYUYMYCAMCMCHTCHGCVYYHYTAHETHTHWYAHYMYMYWY	731			
Qy	125	ACGAATCTGCCCAAATATAATATAAATAACCATATACAGCATACAGTCAAAAGTC	184			
Db	732	YMWYCTACTVEYHHHHYHAYHTTWANAHAMWHEHAYAAABAAWATTHHYHTT	751			
Qy	185	ATACCTGTATGATAGATGACTTTTTTATGTCTCTCTCAAAATACACTTAAATGAG	244			
Db	792	HYMHHTVYUYUYTCGYMCTYHCWYHYHTAYTCWTWTHHMWVTTTTHYHTTWHHTTTH	851			
Qy	245	CATCTCAACCCAGTGAATCCTAGCAGTTCATCGCTCAGCGGTGAACAAATAAATACAT	304			
Db	852	WAWHHTHWCVWWHATTTWATHCWACNTMWEHHEHHEHMHCHAHHTHMCCHH	911			
Qy	305	TACAATTTACGGAAAAAGTGTATATAAAGAGAGACTACAAATTAAGCGCTATAAAC	364			
Db	912	MHCTCHHTHTYHMTCHWMMHMHWEHWMATWNTTWTMMWCMCHHCHMYHMMH	971			
Qy	365	AACATATGTCAAAAGTCTAACTGCCATCCGGCTAACACTTAACCTCAGTCACTTTTA	424			
Db	972	YWCCHYCTCHPATTHYHYMCTCYHYCTWHTTWTAYWAWTAAHMTTATWWWMEKWA	1031			
Qy	425	ACTGTAATAAAACCGGCTTCAGGTGCAAGTCTATATTTATATAT	470			
Db	1032	HWAATWWWWATAWAACTCHHTWYHTHTCTWYHHHTYHMMWAW	1071			

RESULT 15

AC030411.1	CD389151	931 bp	linear	EST 30-MAY-2003
LOCUS	CD389151			
DEFINITION	AGENCOURT_14305279 NIH MGC_173			Homo sapiens cDNA 5', mRNA sequence.
ACCESSION	CD389151			
VERSION	CD389151.1			GI:31227751
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			

ORGANISM

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutharia; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1. (bases 1 to 931)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Daniela S. Gerhard, Ph.D. Office of Cancer Genomics National Cancer Institute / NIH Bldg. 31 Rm10A07 Bethesda, MD 20892 Email: cgabps@email.nih.gov Tissue Procurement: Dr. Jamie Thompson, University of WI

cDNA Library Preparation: Gina Zastrow-Hayes
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LJNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LJNL at:

<http://image.llnl.gov>
 Plate: NDKM70 row: h column: 16
 High quality sequence start: 7
 High quality sequence stop: 617.
 Location/Qualifiers

FEATURES

source

ORIGIN

	Query Match	4.4%;	Score 47.6;	DB 14;	Length 931;
	Best Local Similarity	39.4%;	Fred. No. 43;		
	Matches 137;	Conservative	0;	Mismatches 211;	Indels 0;
	Gaps	0;			
Qy	290	AAACAAATATAACATTAACAATTTACGGAAAAAGACAGCTTAAATATAAAGACAGCTACAAA	349		
Db	545	AAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	604		
Qy	350	TTAAAGCCTATAACCAATATTATGTTCAAAAGTGTTAACTGCCCTCCGGCTAAACACTTA	409		
Db	605	AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	664		
Qy	410	ACTCAGCTCATTTTAACTGTAAATAAAACGGCGTTTCAGGTGCAAGTTATATCTATATA	469		
Db	665	NCCACCTTCTNTTATTATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	724		
Qy	470	TTCCCTGGCGGCACTAAAAAATTGCTTTTGGTGGTATCTGGGATCTACTCTCGAAGT	529		
Db	725	TTNNNNAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	784		
Qy	530	TAACAGTAAAAAGCAGTATAGTCAGACCTATGGAACTTACACTATATAAATATCATTA	589		
Db	785	TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	844		
Qy	590	AATTAACTGATPAGGGAATATTTCAGATATGGTTTACCTCAGTTCAAAA	637		
Db	845	TANTANNNTNNNANTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	892		

Search completed: May 8, 2004, 12:58:25
Job time : 3225 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2004, 08:42:24 ; Search time 98 Seconds
(without alignments)
6149.762 Million cell updates/sec

Title: US-09-839-894-9
Perfect score: 1086
Sequence: 1 atgataagattttattat.....caagtagtcaaacactctag 1086

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A COMB.seq:
2: /cgn2_6/prodata/2/ina/5B COMB.seq:
3: /cgn2_6/prodata/2/ina/6A COMB.seq:
4: /cgn2_6/prodata/2/ina/6B COMB.seq:
5: /cgn2_6/prodata/2/ina/PCUS COMB.seq:
6: /cgn2_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	298.8	27.5	5798	2	US-08-483-101-1
2	43	4.0	832	4	US-09-621-976-2813
3	39.8	3.7	74962	4	US-09-685-853A-3
4	39	3.6	10540	4	US-09-417-4850-5
5	38.8	3.6	342	4	US-09-601-198-115
6	38.4	3.5	4171	4	US-08-956-1715-469
7	38.2	3.5	473	1	US-08-764-100-16
8	38.2	3.5	4970	1	US-08-764-100-14
9	38.2	3.5	4970	1	US-08-764-100-20
10	38.2	3.5	1864976	4	US-08-916-421B-1
11	37.8	3.5	832	4	US-09-621-976-2813
12	37.6	3.5	5562	4	US-10-204-708-63
13	37.4	3.4	533	4	US-09-495-050A-48
14	37.2	3.4	7075	4	US-08-956-1715-263
15	36.6	3.4	2671	6	5168051-9
16	36.4	3.4	1020	4	US-09-463-962-1
17	36	3.3	540	4	US-09-404-879A-2
18	36	3.3	540	4	US-09-338-933-2
19	36	3.3	540	4	US-09-215-681-2
20	36	3.3	540	4	US-09-216-003A-2
21	36	3.3	551	4	US-09-404-879A-137
22	36	3.3	551	4	US-09-338-933-137
23	36	3.3	551	4	US-09-215-681-137
24	36	3.3	551	4	US-09-216-003A-137
25	36	3.3	15016	4	US-09-601-198-60
26	36	3.3	319608	4	US-09-539-333D-1
27	36	3.3	319608	4	US-09-679-409-1

C 28	35.6	3.3	201	4	US-08-956-171E-2594	Sequence 2594, Ap
C 29	35.6	3.3	1689	1	US-07-991-867B-41	Sequence 41, Appl
C 30	35.6	3.3	1689	2	US-08-544-332-41	Sequence 41, Appl
C 31	35.6	3.3	1689	4	US-09-370-861A-41	Sequence 41, Appl
C 32	35.6	3.3	1947	4	US-09-370-861A-74	Sequence 74, Appl
C 33	35.6	3.3	8457	1	US-07-991-867B-1	Sequence 1, Appl
C 34	35.6	3.3	8457	2	US-08-544-332-1	Sequence 1, Appl
C 35	35.6	3.3	8457	4	US-09-370-861A-1	Sequence 1, Appl
C 36	35.6	3.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 37	35.4	3.3	2826	4	US-08-624-655A-1	Sequence 1, Appl
C 38	35.4	3.3	2985	4	US-09-833-381-1138	Sequence 1138, Ap
C 39	35.4	3.3	162450	4	US-09-345-882-1	Sequence 1, Appl
C 40	35.4	3.3	169998	4	US-09-676-610B-24	Sequence 24, Appl
C 41	35.4	3.3	197496	4	US-09-877-177A-1C	Sequence 10, Appl
C 42	35	3.2	1491	4	US-09-134-001C-866	Sequence 866, App
C 43	35	3.2	11770	4	US-08-961-527-172	Sequence 172, App
C 44	35	3.2	640681	4	US-09-790-988-1	Sequence 1, Appl
C 45	34.8	3.2	337	2	US-09-032-684-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-483-101-1
; Sequence 1, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33875
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5798 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
; MOLECULE TYPE: DNA (Genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 499..1215
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: 499..552

```

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 553..1212
FEATURE:
NAME/KEY: CDS
LOCATION: 1255..1767
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1255..1323
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1324..1764
FEATURE:
NAME/KEY: CDS
LOCATION: 1836..4436
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1836..1913
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1914..4433
FEATURE:
NAME/KEY: CDS
LOCATION: 4451..5545
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 4451..4504
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4505..5542

```

US-08-483-101-1

Query Match 27.5%; Score 298.8; DB 2; Length 5798;

Best Local Similarity 58.5%; Pred. No. 1.2e-70;

Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

```

Qy 127 GAATCTCCCCCAACATAATATATAAATACCAATATACAGCATACAGTGAAGTCAT 186
Db 4583 GCATCGTATCTGCTCATTAATATATATATATATATATATATATATATATATAT 4642
Qy 187 ACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 246
Db 4643 TCTCTTTTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4702
Qy 247 TGTCACCAACAGTGAAPCCTAGCAGTTCATCGGTGAGCGGTGAACAAATATAACATTA 306
Db 4703 TGTCGACGAGGAGAAACTCCA-----AATCCTCTCAAGGGGAGACTAATATTAAAGCTA 4756
Qy 307 CAATTACGGAAAGAGACTTTAATAAAGAGAGCTACAAATTAAGGCTATAACAA 366
Db 4757 ATATTACTGAAAGAAAGAGCTGCGCCAGAAACATTAATTAAGGATATAAGAGA 4816
Qy 367 TTATTGTTCAAAAGTGTAACTGC-----CCATCGGCGCTAAACACTTAACTCAGCT 417
Db 4817 TTTTATATGATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4876
Qy 418 CATTTTAATCTTAATAAAGACCGGCTTCAGGTGCAAGTTTATATATATATATCTTGCT 477
Db 4877 ACTGTTAAATGCTGAGGTTTCAATCAAGAGAGTAGATTTCACTTTATATATCCCA 4936
Qy 478 GCGCACTAAATAATTCCTTTTGGTATCTGGGATCTGGGATCTGATGATGATGATG 537
Db 4937 GGTGAAATCTGATGGCTTTCTAACTCGAGCTATATGGGAGGCAACTAGATGATG 4996
Qy 538 AAAAGACCATATAGTGAGACCTATGCACTTACATATATAATATATATATATATTA 597
Db 4997 AAAAGCATATACCATATATATATATATATATATATATATATATATATATATAT 5056
Qy 598 GATAAGGAAATATATCAGATATGTTTACCTCAGTTTCAAAAGTCAAGCTCGGCTG 657
Db 5057 GACAAAGGAAATATATCAGTCTCGGACACCAAGATTTTATAGCATCTTCAAGAT 5116
Qy 658 AACCTGCGTCCNACTGGTGGGGGACATATATATGCAAGAAATCTGATATGCTTT 717

```

```

Db 5117 AATTACGTCCTGAAGTAATGTAAATATATCTGGTAGTAACGTGCTGAGATGTCTC 5176
Qy 718 TATGATGGATATAGTACTAACAGCAGCTCTTTGGAGATAAGATTTTCAGGATAACAATCCT 777
Db 5177 TATGATGGCTATAGTACACATAGTCAAAGTATAGAAATGAGGTTTCAGGATGACTCAA 5236
Qy 778 AATCTGATCGGAAATTTTATCTAAGGAAATTAATATGATGACACCAAGAAATGATAT 837
Db 5237 ACAGGAAATATGAATATAATCTTATAAAACTCGAGAGCCATTAATAAAATTCGCATAT 5296
Qy 838 ACITTTGTCAGTCTCTTTGGGGGTAAAGTTTAACTCCAAACAATGAACTCATTAAT 897
Db 5297 AACTTTCTCTCTTTTAGAGGAGCAGAGCTTTATCCAAATATATGAGGAGCTTTACT 5356
Qy 898 ATTGCTGACGAGCTTCTCTGGAACAACAACCTGGAATACAAATACAGCTGTCAACATCCA 957
Db 5357 ATTAATGATACTTCGTCATTTTATAAAGTGAATCGTATTAAGTCTGTATCTTACCA 5416
Qy 958 GAAATCAGTCTTCGGGTGTTGTTGGCTGGAGCTTTGCAATGGATGCAAAAGTGA 1017
Db 5417 CAGATTAGTATTCAGTACTATGCTGGCCAGCAACTTGACATTTATSTCAGAGCTAAT 5476
Qy 1018 AATCCCGAGGCTGACAAATATATGCGTAATATTAATGTTTACTTTCACACCAAGTAGCAA 1077
Db 5477 AATCCAGAGCGGTGATGATTCAGGAATACTTAACTGTAATCTCTCTAGTAGTTCA 5536
Qy 1078 ACACCTTA 1085
Db 5537 AGTCCGTA 5544

```

RESULT 2

US-09-621-976-2813

; Sequence 2813, Application US/09621976

; Patent No. 5639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET 0542P2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 2813

; LENGTH: 832

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 235..399

US-09-621-976-2813

Query Match

4.0%; Score 43; DB 4; Length 832;

Best Local Similarity 16.8%; Pred. No. 0.034;

Matches 75; Conservative 168; Mismatches 203; Indels 1; Gaps 1;

```

Qy 450 TATTATATATCTCTGCTGGCGAACTAAATAATTTGCTTTTGGTGTATCTGGATGCT 519
Db 1 YRWYWKYTTWYAKCTWKWWSYWKYWKYWKYWKYWKYWKYWKYWKYWKYWKYWKYWKY 60
Qy 520 ACTCTGAAGTTAGA-GTAAAAAGACGATATAGTAGACCTATGGAACCTATACATAAAA 578
Db 61 RYAVNGVTYKKKAMCRTRTKKKKKKKYMMWYWGWRSSYMAWTRTWGTGYAYRSMYTW 120
Qy 579 TATCAGTATTAATTAATCTGATAGGGAATATATCAGATATGTTTACCTCAGTTCAAAG 638
Db 121 YRCWKKKAYKTKTCYSKSGMTWTKWKKGATTTWKKTYWAAIRYWWKMTKRWKRA 180
Qy 639 TGAAGCTCGCGCTGATCTTAACCTGCGTCCAACTGGTGGGGGACATATATGGAAGAA 698

```


Db 181 WYCWVWGWKRWKRSVSARSASRACCCYSSWGSXKMYKRWWRWGWATGAGM 240

QY 699 TTCTGTCGATGTCGTTTATGATGATATAGTACTAAGACAGAGCTCTTTGGAGATAG 758

Db 241 XAWRASOMRKYAGSKSYKSMYCWTRSWKYCYTKARWTGYCYRKGGMWKGSGRWY 300

QY 759 ATTTCAGGATACACCTCCTAAATCTCATGGGAAATTTATCAAGGAAATATAATGATCA 818

Db 301 ASKKYWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRWKRW 360

QY 819 CACCAAGAAATGATATACACTTGTCTCTCTGCGGCTAAAGTTTAACTCCAAC 878

Db 361 RRACAAAATATAATTAATGATGATACATCTCTGCTTATGCAAACTCGGATTAGTTC 420

QY 879 AAATGGAACGCTCAATTAATATTCTGA 905

Db 421 ATAGTCAAGTCAGTTAAATTTCTTA 447

RESULT 3

US-09-685-853A-3/c

Sequence 3, Application US/69685853A

Patent No. 6479270

GENERAL INFORMATION:

APPLICANT: WEI, Ming-hui et al.

TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,

TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,

TITLE OF INVENTION: AND USES THEREOF

FILE REFERENCE: CL000871

CURRENT APPLICATION NUMBER: US/09/685,853A

CURRENT FILING DATE: 2002-05-06

PRIOR APPLICATION NUMBER: 60/182,194

PRIOR FILING DATE: 2000-02-14

NUMBER OF SEQ ID NOS: 3

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 74962

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(74962)

OTHER INFORMATION: n = A,T,C or G

US-09-685-853A-3

Query Match 3.7%; Score 39.8; DB 4; Length 74962;

Best Local Similarity 46.8%; Pred. No. 1.3;

Mismatches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 105 TTTTGGCCCGCGTCACAGGACGAACTCTCCCGAACATATATATTAATACCATAT 164

Db 43211 TTCTAGCTGCTGACAGAGCGAGACTCTGCTCAAAAAAATAAAAAAATAAAAAA 43152

QY 165 TACGACATACAGTGAAGTCTACTCTCTATGATAGGATGACTTTTATCTTCTTC 224

Db 43151 ATCAGTGTGTTGGGATATCTCTTTTAAATTAAGTCTTTTAAATTAAGTCTTAA 43092

QY 225 TCACAAATACCTTAATGGAGCATGCTCCACAGGATGAGATCTTACAGCTTCATCGGTGAG 284

Db 43091 AAAAAAATACCTCAAAAGTAGAGAGTCAAGCTTCCATATCAAGAAATTCATACCTGTG 43032

QY 285 CGGTGAACAAATATATACATTAATTAAGGAAAAAGAGAGTCTTAATAAAGAGAGCT 344

Db 43031 GATTCAACCGCAACAAATAAAAAATTAACAAAAAATTAAGGAAAAAATTAACACTTAA 42972

QY 345 ACAATTAAGGCTTAACATTAAT 372

Db 42971 ATAATAAATAAATAATCAATTAAT 42945

RESULT 4

US-09-417-485D-5

Sequence 5, Application US/09417485D

Patent No. 6541202

GENERAL INFORMATION:

APPLICANT: Long, David M.

APPLICANT: Metz, Anneke M.

APPLICANT: Love, Ruschelle A.

TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes

FILE REFERENCE: 47114-5009-US

CURRENT APPLICATION NUMBER: US/09/417,485D

CURRENT FILING DATE: 2002-06-14

NUMBER OF SEQ ID NOS: 49

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 10640

TYPE: DNA

ORGANISM: Plasmodium falciparum

FEATURE:

NAME/KEY: CDS

LOCATION: (834)..(7385)

OTHER INFORMATION: TERT gene

FEATURE:

NAME/KEY: unsure

LOCATION: (1821)..(1837)

OTHER INFORMATION: m at position 1821 = a or c; w at position 1837 =

OTHER INFORMATION: a or t. Xaa (amino acid) at position 33C = Leu or

OTHER INFORMATION: Ile; Xaa at position 33S = Asp or Gly.

US-09-417-485D-5

Query Match 3.6%; Score 39; DB 4; Length 10640;

Best Local Similarity 45.8%; Pred. No. 1;

Mismatches 135; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 84 AGATGAAGCATAACTAATATTTTGGCCGCGTGCAGGAGCAATCTTCCGCCAACA 143

Db 3767 AATAAAGTAAACAATAATTAACATAACAAGAGACAAATAATTAATAACAATA 3826

QY 144 TAATATATTAATAACCATATTTACGATACAGTGAAGTCTGATGATGATGAT 203

Db 3827 TAATATTAATAACAATAATTAATAACAATAATTAATAATAATAATAACA 3886

QY 204 GACTTTTATGTTGTTCTCTCAATACACTTAAATGGAGCATGTCCACAGTGAGAA 263

Db 3887 CAATGTAACTATCAAAATTCGAGAGACATATTTGATGATTCATTAATAATAACA 3946

QY 264 TCCTAGCAGTTCATCGCTCAGCGGTGAAACAAATATATACATTAACATTAACG 323

Db 3947 GGCTAAATAATTCAGAGACATATTTGATGATTCATTAATAATAACA 4006

QY 324 AAGTTTAAATAAGAGAGCTCAATTAAGGCTATTAACATTAATTTGTTCAA 378

Db 4097 AATATACATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4061

RESULT 5

US-09-601-138-115/c

Sequence 115, Application US/09601198

Patent No. 6531583

GENERAL INFORMATION:

APPLICANT: Cassell, Gail H.

APPLICANT: Chen, Ellison Y.

APPLICANT: Glass, Jennifer S.

APPLICANT: Glass, John I.

APPLICANT: Heiner, Cheryl R.

APPLICANT: Lefkowitz, Elliot

TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA

TITLE OF INVENTION: UREA PLASMA

FILE REFERENCE: UAB-13452/22

CURRENT APPLICATION NUMBER: US/09/601,198

CURRENT FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/073,189

PRIOR FILING DATE: 1998-01-30

NUMBER OF SEQ ID NOS: 181

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 115

```
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-115

Query Match      3.6%; Score 38.8; DB 4; Length 342;
Best Local Similarity 51.8%; Pred. No. 0.33; Mismatches 82; Indels 0; Gaps 0;
Matches 88; Conservative

QY 684 ATATATTGGAAGAAATTCGTTGATATGTCGTTTATGATGATATAGTACTACTACAGCAG 743
DB 303 ATAAATTCGAATATAAATTCGATATATGCTTCTTAATGATTTAGAAATTCCTCAATTAA 244
QY 744 CTCCTTGGAGATAGATTCAGGATACAAATCCTAAATCTGATGGAATTTATCTTAAG 803
DB 243 ATCATCTTAGTAAAGTTAAATTTAAGTACGTAATAATGTAATTTATTTTAAAAAT 184
QY 804 GAAATTAATGATGACACCAAGAAATTCATATATCTTGTCACTTCTCT 853
DB 183 GAATAAAGTGGTCTATATTTTACTTCGTAAATGTTTACCTTTT 134

RESULT 6
US-08-956-171E-469
; Sequence 469, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB24021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 469:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 469:
US-08-956-171E-469

Query Match      3.5%; Score 38.4; DB 4; Length 4171;
```

```
Best Local Similarity 54.2%; Pred. No. 1.1;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 803 GGAAATTAATGATGACACCAAGAAATTCATTAATGATTAATTTCTTCTTGGCGGTA 862
DB 3932 GTACAAATAGAAACACAAATACAAACCCCAAAATATGATATATTTTTCAGMTA 3991
QY 863 AAGTTTAACTCCACAAATGGAAGTCATTAATTAATTTGCTGACGAGTTCTCTGGAAA 922
DB 3992 ACAGTCGAATAAGACCAAAAGTAAAGGATATTAATCCTACAAAATAGTTGCCATCACA 4051
QY 923 CAAGTCGAATAGATTACAGCTG 946
DB 4052 TAAATAGATAGCGCTATATATG 4075

RESULT 7
US-08-764-100-16
; Sequence 16, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gieles L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sandoz Agro, Inc
; STREET: 975 California Avenue
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 80C
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,450
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-764-100-16

Query Match      3.5%; Score 38.2; DB 1; Length 473;
Best Local Similarity 43.8%; Pred. No. 0.53;
Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 420 TTTTACTCTAATAAACAACGGCTTCAGTGCAGTTATATATATATATATATATATCTCTG 479
```

```

Db 47 TTGAATGTTTAAAGTAAATAAAGCAAGATAAATAAATACTATATATATATATATATATAGA 106
Qy 480 CGAAGTAAATAATTTGCTTTTGGTGTATCTGGATGCTACTGAGTTAAGATGAAA 539
Db 107 AGTATAAATAATATATGATTTTGTGTTTAAATAAATAAATAAATAAATAAATAAATAA 166
Qy 540 AAGACGATATAGTGAAGCTATGAACCTTACACTATATAATATCTATTAATAAATGA 599
Db 167 AAAAAATAAACAATAAACAATAAACAATAAACAATAAACAATAAACAATAAATAAATAA 226
Qy 600 TAAGGGAATAATTCAGATATAGTTCAGTCTCAAAAGTCAAGCTCGCGTCGATCTTAA 659
Db 227 AAGTTGAAAAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 286
Qy 660 CTTCGCTCAACTGGTGGGGSCATATATGGAGAAATCTCTGATATATGCTTTTA 719
Db 287 TCTAAGACCTTTTATTTGTTTATATACATTTTATTTGTTTGTGTTTATTTTATTTTA 346
Qy 720 TGATGGATATAGTACTAACAGCAGCTCTTTGGAGATAAGATTCAGGATACAATCTTAA 779
Db 347 TTATTTTATATTTTATATAGTTTGTCTTATTTAAGCTTATTTAGACAAATTAATTT 406
Qy 780 ATCTGATGGGAATTTTAT 798
Db 407 ATTGATTACATCTCT 425

```

RESULT 8

```

US-08-764-100-14
; Sequence 14, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; STREET: Sandoz Agro, Inc
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9
; FILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700r1s, Ailer E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4970 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-764-100-14

Query Match 3.58; Score 38.2; DB 1; Length 4970;
Best Local Similarity 43.88; Pred. No. 1.3; Mismatches 0; Gaps 0;
Matches 66; Conservative 0; Indels 213;

Qy 420 TTTTAACTGTAATAAAAAACGGGCTTCAGGTGCAAGTTTATATTTATATATATCTGCTGG 479
Db 1014 TTGAATGCTTTAAGTAATAAATAAAGCAAGATAAATAAATACTATATATATATATAGA 1103
Qy 480 CGAAGTAAATAATTTGCTTTTGGTGTATCTGGATGCTACTGAGTTAAGATGAAA 539
Db 1104 AGTATAAATAATATATGATTTTGTGTTTAAATAAATAAATAAATAAATAAATAA 1163
Qy 540 AAGACGATATAGTGAAGCTTATGAACCTTACACTATATAATATCTATTAATAAATGA 599
Db 1164 AAAAAATAAACAATAAACAATAAACAATAAACAATAAACAATAAATAAATAAATAA 1223
Qy 600 TAAGGGAATAATTCAGATATAGTTCAGTCTCAAAAGTCAAGCTCGCGTCGATCTTAA 659
Db 1224 AAGTTGAAAAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1283
Qy 660 CTTCGCTCAACTGGTGGGGSCATATATGGAGAAATCTCTGATATATGCTTTTA 719
Db 1284 TCTAAGACCTTTTATTTGTTTATATACATTTTATTTGTTTGTGTTTATTTTATTTTA 1343
Qy 720 TGATGGATATAGTACTAACAGCAGCTCTTTGGAGATAAGATTCAGGATACAATCTTAA 779
Db 1344 TTATTTTATATTTTATATAGTTTGTCTTATTTAAGCTTATTTAGACAAATTAATTT 1403
Qy 780 ATCTGATGGGAATTTTAT 798
Db 1404 ATTGATTACATCTCT 1422

```

RESULT 9

```

US-08-764-100-20/c
; Sequence 20, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.
; APPLICANT: Peters, Dirk
; APPLICANT: Goldbach, Robert W.
; TITLE OF INVENTION: Improvements in or Relating to Organic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; STREET: Sandoz Agro, Inc
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,100
; FILING DATE: 06-DEC-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,064
; FILING DATE:
; APPLICATION NUMBER: US 08/032,235
; FILING DATE: 17-MAR-1993
; APPLICATION NUMBER: GB 9206016.9

```

```
;
; PILING DATE: 19-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5773700ris, Allen E.
; REGISTRATION NUMBER: 34,490
; REFERENCE/DOCKET NUMBER: 137-1061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 354-3592
; TELEFAX: (415) 857-1125
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4970 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-764-100-20

Query Match          3.5%; Score 38.2; DB 1; Length 4970;
Best Local Similarity 43.8%; Pred No. 1.3;
Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

Qy 420 TTTTAACTGTTAATAAAACGGCGGTCAGGTGCAAGTTTATATTATATATATTCCTGCTGG 479
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3927 TTGAATGTTTAAATAAAATAAAGCAAGATAAAAACCTATATATATATATATATAGA 3868
Qy 480 CGAACTAAATAATTCCTTTTGGTGGTATCTGGGATCTCTCTGAGTTAAGAGTTAA 539
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3867 AGTATAAATATATATGATTTGTTTAAATAAACAATCAAAAACCAAAAGAAAAAG 3808
Qy 540 AAGACGATATAGTGAGACTTGGAACTTACACTATAATATCACTATAAATTAACCTGA 599
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3807 AAAAAATAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAATAGAAA 3748
Qy 600 TAAGGAATATTCAGATATATGTTACCTCAGTCAAAAAGTCAGCTCCGTCGATCTTAA 659
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3747 AAGTTGAAAAAACCACAAAAATTTTGTGTAATAAATAGGCTCCGGCCAGATTGG 3688
Qy 660 CTTCGCTCAACTGTGGGGCAGCATATATGGAAGAATCTGTGTGATATGCTTTTA 719
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3687 TCTAAGACCTTTTATTTGCTTTTATACATTTTATTTGTTTGTGTTTATTTT 3628
Qy 720 TGATGCATATAGTAAACAGCAGCTCTTTGAGATAAGATTTCAGGATAACAATCTTAA 779
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3627 TTATTTTATATTTTATATATAGTTTGCCTTATTTAACACTTATTTAGACAAATTAATTT 3568
Qy 780 ATCGATCGGAATTTAT 798
Dy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dy 3567 ATTGATTACAATCAATCT 3549

RESULT 10
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Jannaschli
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/524,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234314)..(234314)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
```

```
; LOCATION: (552241)..(552241)
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084930)..(1084930)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
```

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match          3.5%; Score 38.2; DB 4; Length 1664976;
Best Local Similarity 51.5%; Pred. No. 11;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 747 TTGGGATGAAGATTTCCAGGATACCAATCCATAATCTGATGGGAAATTTATCTAAGAA 806
Db 1486909 TGTGAATTTGAATTAAGAAACAAATCCTGTTTATGTTGAAAAATCCTATGAAAAACT 1486968

Qy 807 AATAAATGATGACACCAAGAAATTCATATATCTTGTCACTTCTCTTGGCGGTAAAG 866
Db 1486969 CTTAGCTGATATAACAAAGATGATGCAATCTATTGAACCTAGGATTTGAGGGGGCTT 1487028

Qy 867 TTTAACTCCAAAGAAATGGAACGTCATTAATAATTTGCTGACGCACTTCTCT 917
Db 1487029 TTTAAATAAAACAGTATATCTTTATTATGAAAAATGAAGAAATCATCT 1487073

RESULT 11
US-09-621-976-2813/c
; Sequence 2813, Application US/09621975
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813
```

```
Query Match          3.5%; Score 37.8; DB 4; Length 832;
Best Local Similarity 13.4%; Pred. No. 0.85;
Matches 49; Conservative 162; Mismatches 159; Indels 3; Gaps 1;

Qy 568 TACACTATAATATCACTATTAAATTAAGTGAAGGGAATATTCAGATATGTTACCT 627
Db 370 TATTTTGTYYWKKTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWYTWY 311

Qy 628 CAGTTCAAAAGTGACGCTCGGTCGATCTTAACCTTCGCTCCAACTCGTGGGGCACATAT 687
Db 310 WYWKRMSTWYCYWCKCMYGRRCATWYTWARGRMVSYAWGKWSMESAYSMCTRY 251

Qy 688 ATTGGAAGAAATCTCTTGTGATATGCTTTTATGAGGATATAGTACTACACAGCTCT 747
Db 250 YRKGSTYWTNKTCTCATWYWKYKRMWSKTCWSGSGGYMTSYTSYSYWSWY 191

Qy 748 TTGGAGATAAGATTTCCAGGATAACATCCTAAATCTGATGGGAAAATTTATCTTAAGGAA 807
Db 190 TMCWWRWSTWYTWYWKWKKWYRZATTTWRAMWWWA---ATWYMWYMWYMWYMWY 134

Qy 808 ATAATGATGACACCAAGAAATTCATATATCTTTGTCACTTCTCTTGGCGGTAAAGT 867
Db 133 AMYRRITMMWGYRWRKYSRTRTCAWAYANKTKGSYWCWWRKWCWMMMAVYG 74
```



```
QY 541 AGACGATATAGTGGAGACCTATGGAACCTTACACTATAAATATCACTATTAATAAATACTGAT 600
DB 4736 AGACGATATAGTGGAGACCTATGGAACCTTACACTATAAATATCACTATTAATAAATACTGAT 4795
QY 601 AAGGGAATATTCAGATATGGTTTACCTCAGTTCAAAAGTCAGGCTCGCTCGATCTTAAC 660
DB 4796 AAGGGAATATTCAGATATGGTTTACCTCAGTTCAAAAGTCAGGCTCGCTCGATCTTAAC 4855
QY 661 TTGGCTTCCAACTGGTGGGGGCACATATATGGAAGAAATTCCTGTCATATGCTTTTAT 720
DB 4856 TTGGCTTCCAACTGGTGGGGGCACATATATGGAAGAAATTCCTGTCATATGCTTTTAT 4915
QY 721 GATGATATAGTAACTAAGCAGCTCTTTGGAGATAGATTTTCAAGATTAACATCTTAA 780
DB 4916 GATGATATAGTAACTAAGCAGCTCTTTGGAGATAGATTTTCAAGATTAACATCTTAA 4975
QY 781 TCTGATGGGAAATTTTATCTAAGGAAATTAATGATGACACCAAAAGAAATTTGCATATCT 840
DB 4976 TCTGATGGGAAATTTTATCTAAGGAAATTAATGATGACACCAAAAGAAATTTGCATATCT 5035
QY 841 TTGTCACTCTCTTGGCGGGTAAAGCTTTAACTCCAAACAAATGGAAACGTCATTAATAT 900
DB 5036 TTGTCACTCTCTTGGCGGGTAAAGCTTTAACTCCAAACAAATGGAAACGTCATTAATAT 5095
QY 901 GCTGAGCAGCTCTCTGGAACACACTGGAATAGAAATACAGCTGTCACCATGCCAGAA 960
DB 5096 GCTGAGCAGCTCTCTGGAACACACTGGAATAGAAATACAGCTGTCACCATGCCAGAA 5155
QY 961 ATCAGTGTTCGGGTGTGTGTGGCTGGACGTTTCAATTTGATGCAAAAGTGGAAAT 1020
DB 5156 ATCAGTGTTCGGGTGTGTGTGGCTGGACGTTTCAATTTGATGCAAAAGTGGAAAT 5215
QY 1021 CCGGAGCTGGACATATATGGTAAATATTAATGTTACTTTACACCAAGTAGTCAACAA 1080
DB 5216 CCGGAGCTGGACATATATGGTAAATATTAATGTTACTTTACACCAAGTAGTCAACAA 5275
QY 1081 CTCTAG 1086
DB 5276 CTCTAG 5281

RESULT 3
US-10-311-455-1564/c
; Sequence 1564, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1564
; LENGTH: 10279
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1564

Query Match 4.5%; Score 49.4; DB 15; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.32;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
```

```
QY 140 AACATAATATTTAATTAACCATATTTACAGCATACAGTCAAGTGCATATCTCTGTATGATA 199
DB 7605 ATCATATTAATAAATAATTTAAATAAATAAACAATAAATAAATAATTAATAATCCAAAAAACA 7546
QY 200 GGATGACATTTTATGTTTGTCTCTCAAAATACACACTTTAATGGAGCATGTCACAAACGAGTG 259
DB 7545 ATAAACCATTTATTAACCTTATTTTTCGAAATAAACCACAAAAAATCTTACTTCTCAATA 7486
QY 260 AGAATCTTAGCAGTTTCATCGGTACGGGGAACAAATAATAACATTTACATTTACGGAAA 319
DB 7485 AAAAAACAATAAATCTTAATTTCAACACGATATAAATAAATAAATAAATAAATAAATAAATAA 7426
QY 320 AAAGAAGTTTATAAATAAGAGAGCTACAAATTTAAAGGCTATAAACAATTTCTTCAAAA 379
DB 7425 ACATTAATAATACATATATATATATCAAAATCCAAACTTTTATCAACAAACATTAATAA 7366
QY 380 GTGTTAACTGCCATCCGGCCCTAACACCTTTAACTCAGCTCATTTTAACTGTAATAAATAAC 438
DB 7365 ACTATATCTTCATAAACCACCAATATTTTATATATCTTCATTTAAATTCATTTAATATAC 7307

RESULT 4
US-10-240-589C-86/c
; Sequence 85, Application US/10240589C
; Publication No. US20040076596A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: DNA repair
; FILE REFERENCE: 5013.1008
; CURRENT APPLICATION NUMBER: US/10/240,589C
; PRIOR FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03972
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 96
; LENGTH: 10279
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-589C-86

Query Match 4.5%; Score 49.4; DB 17; Length 10279;
Best Local Similarity 47.8%; Pred. No. 0.32;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
```

```
QY 380 GTGTTAACTGCCATCCGGCCCTAACACTTAAGCTACGCTCATTTTAACTGTAATAAAAC 438
Db 7365 ACTATATCTTCTAAACACCCCAATATTTATATATCTCTGATTAAATTCATTABATAC 7307

RESULT 5
US-10-240-453-281/c
; Sequence 281, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BEBLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 281
; LENGTH: 9504
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-281

Query Match 4.4%; Score 48.2; DB 15; Length 9504;
Best Local Similarity 46.8%; Pred. No. 0.61;
Matches 152; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 317 AAAAAGAGTTTAAATAAAAGAGAGCTACAAAATTAAAGGCTATAAAACAATATTGTTC 376
Db 9072 AACATAAATTTCTTAAATAATAAAAAATTAACTTAAATATAAAAAAATAAAATAAA 9013

QY 377 AAAGTGTTAACTGCCCATCCGGCCCTAACACTTAAGCTACGCTCATTTTAACTGTAATA 436
Db 9012 AAATTAACGACAAAAAACAATCTTAACAATTAATAAATAAATAAATAAATAAATAA 8953

QY 437 ACGCGGCTTCAGTGCAAGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 496
Db 8952 AAACAAAAAGTTTAAATAATTTATATAAATAAATAAATAAATAAATAAATAAATAA 8893

QY 497 CTTTGGTGGTATCTGGGATGCTACTCTGAAGCTTAAGAGTAAAGAGCAATTAAGTGA 556
Db 8892 TTATAAAAAACAATCACTTAATAAATCAATAATTTAAATTTATCTTTTAAACACTAAA 8833

QY 557 CCTATGGAAGCTTACACTATAAATATACATTAATTAATTAAGTGAAGGGAATATTCAGA 616
Db 8832 AATCACTAAAAAATAAATCAATTTCTTAAACAATAAATAAATAAATAAATAAATAA 8773

QY 617 TATGGTTACTCTAGTCAAAAGTGA 641
Db 8772 AATATATTTCTTATTTCTTAAATAAA 8748

RESULT 6
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/80
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match 4.4%; Score 48.2; DB 15; Length 3673778;
Best Local Similarity 46.8%; Pred. No. 13;
Matches 152; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 146 ATATATTAATAAACCATATTACAGCATACAGTGAAGTATATCTCTGTATGATAGATGA 205
Db 2363757 ATATATTTAACACACACATACCTAACTAATTTTATATTTTAAATAAATAAATTC 2363698

QY 206 CTTTTTATGTTTGTCTTCTCAATATACACTTAAATCGAGCATGTCACACAGTGAGATC 265
Db 2363697 AATTTAAATTAATACTAATATCGAACTCCTAATAATCTACCTCTTACTTCCCAAATA 2363638

QY 266 CTAGCAGTTCATCGCTCAGGGTGAACAAATATAACATTACAAATTTACGGAAAAAGAA 325
Db 2363637 CTAAATTAACAACGTAATATCGACAAACAATACTACATTTTAAATAATAAATAA 2363578

QY 326 GTTTAATAAAAGAGAGAGCTACAAAATTAAAGGCTATAAACAATTTATTTTCAAAAGTGA 385
Db 2363577 CTTCAAAAAATAAATAATTTTAAATCTATACAAAAAATAAATAAATAAATAAATAA 2363518

QY 386 ACTGCCCATCCGGCTTAACACTTAACCTCAGCTCAATTTTAACTGTAAATAAACAACGCGCTT 445
Db 2363517 ATATATAATAAATAAATAAATAATTTTTCNTTAAATTTTAAAAAATAATAATCTTAAATA 2363458

QY 446 CAGGTGCAAGTTTATATTTATATATAT 470
Db 2363457 ATATATAATTAATAAATACTTATATAT 2363433
```

```
RESULT 7
US-10-424-599-7070
; Sequence 7070, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 7070
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT3847_106392C.1
US-10-424-599-7070
```

```
Query Match 4.1%; Score 44; DB 13; Length 628;
Best Local Similarity 46.4%; Pred. No. 1.6;
Matches 143; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
```


APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287979
LENGTH: 499
TYPE: DNA
ORGANISM: Human
US-10-027-632-287979

Query Match 4.0%; Score 43.2; DB 16; Length 499;
Best Local Similarity 48.0%; Pred. No. 2.3;
Matches 120; Conservative 1; Mismatches 129; Indels 0; Gaps 0;
QY 669 AACTGCTGGGGGCACATATATGGAGAAATCTCTGTGATATGTCCTTTATGATGGATA 728
DB 306 AACATTCGAGGAAGATAAATGACAAATGCTTTTAAATAGTATTAATGTTATATAA 247
QY 729 TAGTACTAACAGCAGCTCTTTGGAGATAAGATTTCCAGGATAACAATCCATAATCTGATGG 788
DB 246 TAGTAGATAGAAAAGCTCACCTTAATTAATATATTAATAATTAACCTAAGCTGTAATTTCT 187
QY 789 GAAATTTTACTAAGGAAATTAATGATGACACCAAGAAATGATATATCTTCTACT 848
DB 186 TGAACCTCAATATGTTTAAACAAAAGCAACAAATATAAATGTTACCATGCCACC 127
QY 849 TCTCTGGCGGTAAAGTTTAACCCAAATGGAACGTCATTAAATATTTGTCGAGCG 908
DB 126 TAGTGGACAGGTCATTTTGATTACACCTCGAATCCACAGCATTCATATGCTTCCAA 67
QY 909 AGCTTCTCTG 918
DB 66 AGTGTCTCTG 57

RESULT 11

US-10-027-632-287980/c
Sequence 287980, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 287980
LENGTH: 499
TYPE: DNA
ORGANISM: Human
US-10-027-632-287980

Query Match 4.0%; Score 43.2; DB 16; Length 499;
Best Local Similarity 48.0%; Pred. No. 2.3;
Matches 120; Conservative 1; Mismatches 129; Indels 0; Gaps 0;
QY 669 AACTGCTGGGGGCACATATATGGAGAAATCTCTGTGATATGTCCTTTATGATGGATA 728
DB 306 AACATTCGAGGAAGATAAATGCTCAATTTAATTAATATTAATAATTAACCTAAGCTGTAATTTCT 247
QY 729 TAGTACTAACAGCAGCTCTTTGGAGATAAGATTTCCAGGATAACAATCCATAATCTGATGG 788
DB 246 TAGTAGATAGAAAAGCTCACCTTTAATTAATATATTAATAATTAACCTAAGCTGTAATTTCT 187
QY 789 GAAATTTTACTAAGGAAATTAATGATGACACCAAGAAATGATATATCTTCTACT 848
DB 186 TGAACCTCAATATGTTTAAACAAAAGCAACAAATATAAATGTTACCATGCCACC 127
QY 849 TCTCTGGCGGTAAAGTTTAACCCAAATGGAACGTCATTAAATATTTGTCGAGCG 908
DB 126 TAGTGGACAGGTCATTTTGATTACACCTCGAATCCACAGCATTCATATGCTTCCAA 67
QY 909 AGCTTCTCTG 918
DB 66 AGTGTCTCTG 57

RESULT 12

US-10-094-240-10/c
Sequence 10, Application US/10094240
Publication No. US20030082637A1
GENERAL INFORMATION:
APPLICANT: ZWIEBEL, LAURENCE J.
TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
FILE REFERENCE: N8289
CURRENT APPLICATION NUMBER: US/10/094,240
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 10/056,405
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: 60/264,649
PRIOR FILING DATE: 2001-01-25
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 4985
TYPE: DNA
ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 3.9%; Score 42.8; DB 15; Length 4985;
Best Local Similarity 41.7%; Pred. No. 9.4; Mismatches 467; Indels 2; Gaps 1;
Matches 335; Conservative 0;
QY 8 AGATTTTATTTTATTTTACATTTGTTTTCTCTTCAGTACTTTTACATTTGCTGTAACGG 67
DB 3149 ACAGTTACTGGAATTTACTATTTACTATTTTACTTTTATTTTATTTTATTTATGGA 3090
QY 68 CAGATAAATTTCCCGGAGATGAAGCAATTAATTTTGGCCCGCTGACAGGACG 127
DB 3089 TTACTGGAAATATTAGAGATAAAAAAACAATAATAAATAAATAAATAAATAAATAA 3030
QY 128 AATCTCCCCCAACATATATATTAATAACCATATTTACAGCATACAGTGAAGTCATA 187

Db 3029 ACAACACAATAAATAGAACACAAACAACAATAAAGAAATTAATTAACAATA 2970
Qy 188 CTCCTGATGATAGGATGACTTTTATGTTCTCTCTCACAATACACTTAATGAGCAT 247
Db 2969 GCAATAAATAGAAATTAATTAACAATAGCAATAAATAAATAAATAAATAAATA 2910
Qy 248 GTCCAAACCAAGTACAGTCTACAGTCTCAT--CGGTACGGGTGAAACAATAATACATT 305
Db 2909 ATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2850
Qy 306 ACAATTACCGGAAAGAGTTTAAATAAAGAGAGCTACAAATTAAGGCTAATAACA 365
Db 2849 AATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2790
Qy 366 ATTATTGTTCAAAAGTGTAACTGCGCCATCCGGCTTAACTAACTAGCTCAATTTAA 425
Db 2789 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2730
Qy 426 CTGTAATAAAGAGCGGCTTCAGGTGCAAGTTTATATTTATATATTTCTCTGCGGAACT 485
Db 2729 ATAATCTATAAATGAATAAAGTAAATAAATAAATAAATAAATAAATAAATAAATA 2670
Qy 486 AAAAAATTTGCGCTTGGTATCTGGGTGCTACTCTGCACTTAAGAGTAAAGAGAGCG 545
Db 2669 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2610
Qy 546 ATATAGTACAGCTTACCACTTACATAAATAAATAAATAAATAAATAAATAAATAAAGG 605
Db 2609 TAATAATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2550
Qy 606 AATAATTCAGATAGTTACTCAGTTCAAAGTACAGCTCGCGTGGATCTTAACCTTGGG 665
Db 2549 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2490
Qy 666 TCCAACTGGTGGGGACATATATTCGAAGAAATTCCTTGATATGCTTTTATGATGG 725
Db 2489 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2430
Qy 726 ATATAGTACAAAGAGCTCTTTGGAGATAAGATTTTCAGATAACAATCCTAAATCTGA 785
Db 2429 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2370
Qy 786 TGGGAAATTTTATCTAAGGAAAT 809
Db 2369 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2346

RESULT 13
US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFACATORY GENES, POLYPEPTIDES, AND METHODS OF
; FILE OF INVENTION: USE THEREOF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

Query Match 3.9%; Score 42.8; DB 15; Length 4985;
Best Local Similarity 41.7%; Pred. No. 9,4;
Matches 335; Conservative 0; Mismatches 467; Indels 2; Gaps 1;

Qy 8 AGATTATTTTATTTTACATTTGTTTTCTCTTCAGTACTCTTTTACATTTGCTGATCGG 67
Db 3149 ACAGTTACTGGATTAATTTACTATTTTACTATTTTATTTTATTTTATTTATGGA 3090
Qy 68 CAGATAAAAATTTCCGGAGATGAAGCATACATTAATTTTGGCCCGGTGACGAGAGCG 127
Db 3089 TTACTGGGAATTAATTAGAGATAAATAAATAAATAAATAAATAAATAAATAAATAAACA 3030
Qy 128 AATCTTCCCGCAACATTAATTAATTAACCATTAATACACATACAGTGAAGTCAATA 187
Db 3029 ACAACAACAATAAATAAAGAACCAACAACAATAAAGAAATTAATAAATAAATAAATA 2970
Qy 183 CTCCTGATGATAGGATGACTTTTATGTTCTCTCTCACAATAACATTAATGAGCAT 247
Db 2969 GCAATAAATAGAAATTAATTAACAATAGCAATAAATAAATAAATAAATAAATAAATA 2910
Qy 243 GTCCAAACCAAGTACAGTCTACAGTCTCAT--CGGTACGGGTGAAACAATAATACATT 305
Db 2909 ATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2850
Qy 306 ACAATTACCGGAAAGAGTTTAAATAAAGAGAGCTACAAATTAAGGCTAATAACA 365
Db 2849 AATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2790
Qy 366 ATTATTGTTCAAAAGTGTAACTGCGCCATCCGGCTTAACTAACTAGCTCAATTTAA 425
Db 2789 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2730
Qy 426 CTGTAATAAAGAGCGGCTTCAGGTGCAAGTTTATATTTATATTTCTCTGCGGAACT 485
Db 2729 ATAATCTATAAATGAATAAAGTAAATAAATAAATAAATAAATAAATAAATAAATA 2670
Qy 486 AAAAAATTTGCGCTTGGTATCTGGGTGCTACTCTGCACTTAAGAGTAAAGAGAGCG 545
Db 2669 AATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2610
Qy 546 ATATAGTACAGCTTACCACTTACATAAATAAATAAATAAATAAATAAATAAATAAAGG 605
Db 2609 TAATAATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2550
Qy 606 AATAATTCAGATAGTTACTCAGTTCAAAGTACAGCTCGCGTGGATCTTAACCTTGGG 665
Db 2549 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2490
Qy 666 TCCAACTGGTGGGGACATATATTCGAAGAAATTCCTTGATATGCTTTTATGATGG 725
Db 2489 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2430
Qy 726 ATATAGTACAAAGAGCTCTTTGGAGATAAGATTTTCAGATAACAATCCTAAATCTGA 785
Db 2429 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2370
Qy 786 TGGGAAATTTTATCTAAGGAAAT 809
Db 2369 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2346

RESULT 14
US-10-240-485-3/c
; Sequence 3, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: CLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; TITLE OF INVENTION: Diagnoses of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/02970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.6

PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 202
SEQ ID NO 3
LENGTH: 6137
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-3

Query Match 3.9%; Score 42.8; DB 15; Length 6137;
Best Local Similarity 51.0%; Pred. No. 10;
Matches 102; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 137 CCACACATATATTAATTAACCATATTACAGCATACAGTGAAGTCATCTCTGTATG 196
DB 426 CCACATTTATTAATTAATTAACCATTAACCACTCTTAATATATCTCTCCCAATCCA 367
QY 197 ATAGGATGACATTTTATGTTTGTCTTCTCAATATACATTAATGGAGCATGTCCACCA 256
DB 366 TTAATTTTATTAATTTCAATTAATTAACATATCATCTATCAATTAATTAATACAA 307
QY 257 GTGAGATCCCTAGCATTCCTCGTCCAGCGGTGAACAAATATACATTAACATTTACGG 316
DB 306 TCTTACAACTCACCTTTTCAATATACCACTTAATAAACAATATCACTAAACTCAAAA 247
QY 317 AAAAAAGAGTTTAAATAA 334
DB 246 AATAAATACTTCAAAA 229

RESULT 15
US-10-312-841-1/c
Sequence 1, Application US/10312841
Publication No. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
FILE REFERENCE: E01/1208/WO
CURRENT APPLICATION NUMBER: US/10/312,841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 1
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (3294164)
US-10-312-841-1

Query Match 3.9%; Score 42.8; DB 15; Length 3673778;
Best Local Similarity 45.1%; Pred. No. 2.5e-02;
Matches 158; Conservative 0; Mismatches 192; Indels 0; Gaps 0;
QY 252 AACCATGAGATCTCTAGCATTCCTCGTCCAGCGGTGAACAAATATACATTAACATTTACATTT 311
DB 3159690 AACCATTCATATCATATCAATATCACTTAACAAACCCCAATACAAAATATTTAA 3159631
QY 312 TACGAAATAAGAGTTTAAATAAAGAGAGCTACAAATTAAAGCTTATAACCAATTTATT 371
DB 3159630 AATAAATCCAAAAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3159571
QY 372 GTCAAAAGTGTAACTCCCGCTACAGCTTAACACTTAACCTCACTTTTAACCTGTAA 431

Search completed: May 8, 2004, 13:10:12
Job time : 532 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:11:11 ; Search time 59 Seconds
(without alignments)
1728.808 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1866
Sequence: 1 MNKILPITLFPSSVLPTFA.....EAGQMGNNVITFPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: geneseqp1980s: *
2: geneseqp1990s: *
3: geneseqp2000s: *
4: geneseqp2001s: *
5: geneseqp2002s: *
6: geneseqp2003as: *
7: geneseqp2003bs: *
8: geneseqp2004s: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1886	100.0	361	5	AAM50343 ETEC CS4
2	1886	100.0	361	6	ABU09014 Csa opero
3	1881	99.7	1715	6	ABU09015 Csa opero
4	927.5	49.2	364	2	AAY22326 Pilin pro
5	275.5	14.6	359	4	AAB45919 S. enteri
6	120	6.4	1335	6	ABU14929 Protein e
7	119.5	6.3	770	3	AAY50814 D. melano
8	117.5	6.2	1386	6	ABB82573 H. influe
9	117.5	6.2	1449	6	ABB82570 H. influe
10	114	6.0	650	2	AAR47575 Alpha-agg
11	113.5	6.0	513	6	ABU33527 Protein e
12	111.5	5.9	1005	3	ABU01833 Haemophil
13	111.5	5.9	1011	3	ABU01832 Haemophil
14	110.5	5.9	1529	2	AAR41732 High mole
15	110.5	5.9	1601	2	AAB30292 Non-type
16	109	5.8	2383	3	AAB15945 E. coli p
17	109	5.8	2383	6	ABU14928 Protein e
18	108	5.7	2529	6	ABU30910 Prote:n e
19	107	5.7	1095	3	ABU01835 Haemophil
20	107	5.7	1101	3	ABU01834 Haemophil
21	106.5	5.6	448	6	ABM70719 Photorhab
22	106	5.6	671	2	AAR85290 Streptoco
23	106	5.6	1222	3	ABU01830 H. influe
24	106	5.6	1228	3	ABU01828 Haemophil
25	105	5.6	598	5	ABB47313 Listeria

26	105	5.6	598	6	ABL32687	Abu32687 Protein e
27	104.5	5.5	1440	5	ABE54801	AbE54801 Lactococc
28	103.5	5.5	2399	6	ABU31130	Abu31130 Protein e
29	102	5.4	444	6	ABM69257	ABM69257 Photorhab
30	101	5.4	1338	2	AAR41731	Aar41731 High mole
31	101	5.4	1598	2	AAM30291	Aam30291 Non-type
32	100.5	5.3	1004	3	AA301841	Aab01841 Haemophil
33	100.5	5.3	1010	3	ABU01840	Aab01840 Haemophil
34	99.5	5.3	969	3	ABU01827	Aab01827 Haemophil
35	99.5	5.3	975	3	ABU01826	Aab01826 Haemophil
36	99.5	5.3	1300	6	ABU39488	Abu39488 Protein e
37	99	5.2	839	5	ABP56002	ABP56002 Chlamydia
38	99	5.2	839	5	ABP98211	ABP98211 Chlamydia
39	99	5.2	839	6	ABU66267	Abu66267 C. psittac
40	99	5.2	932	3	ABU01843	Aab01843 Haemophil
41	99	5.2	998	3	ABU01842	Aab01842 Haemophil
42	99	5.2	1948	6	ADA09347	Ada09347 Haemophil
43	97	5.1	418	3	AAG30907	Aag30907 Arabidops
44	97	5.1	454	3	AAG30906	Aag30906 Arabidops
45	97	5.1	957	3	AAE01839	Aab01839 Haemophil

ALIGNMENTS

RESULT 1
AAM50343 standard; protein; 361 AA.
XX
AC AAM50343;
DT 18-FEB-2002 (first entry)
XX
DE ETEC CS4 pilus CsaE tip associated protein.
XX
KW CS4 pilus; enterotoxigenic; ETEC; csa operon; CsaA; fimbrial; vaccine;
KW diarrhoea; antibacterial; anti-diarrheic.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= signal_peptide 24..361
FT Protein /label= Mature_protein
XX
PN WO200181582-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012914.
XX
PR 20-APR-2000; 2000US-0198686P.
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI Altboum Z, Levine MM, Barry EM;
XX
XX WPI; 2002-049280/06.
XX N-PSDB; AAI70763, AAI70780.
XX
PT New nucleotide sequence, useful as immunogenic agent for generating
PT immune response against recombinant product of the operon, comprises csa
PT operon which encodes enterotoxigenic Escherichia coli-CS4 pill.
XX
PS Claim 10; Page 59; 81pp; English.
XX
CC The present sequence is that of the tip associated protein CsaE of
CC enterotoxigenic Escherichia coli (ETEC) strain E11891A. CsaE is encoded
CC by the csaE gene (see AAI70763) of the E. coli E11891A csa operon. This
CC operon includes 5 contiguous genes, csaA-csaE, which encode the synthesis
CC of ETEC-CS4 pill. It has been expressed in attenuated Shigella strain
CC of ETEC-CS4 pill, constructing the Shigella expressing CS4 fimbriae vaccine
CC CVD1204 guaba, constructing the Shigella expressing CS4 fimbriae vaccine

CC strain CVD1204 (pCA2-CS4). The CsaE protein has a calculated mol.wt. of
 CC 40102.4 and a theoretical pI of 8.74. It shows homology to similar
 CC proteins from other ETEC fimbriae. Recombinant CsaA-CsaE polypeptides are
 CC used in claimed immunogenic compositions to generate an immune response
 CC in a subject. these prevent ETEC colonisation, and hence protect against
 CC diarrhoea
 XX
 SQ Sequence 361 AA;
 Query Match 100.0%; Score 1886; DB 5; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.6e-168;
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
 DB 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
 QY 61 SHLYDRMTFLCLSSHTNLNGACPTSENPSSSSVSGETNITLQFTKRSILKRELQIKGY 120
 DB 61 SHLYDRMTFLCLSSHTNLNGACPTSENPSSSSVSGETNITLQFTKRSILKRELQIKGY 120
 QY 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180
 DB 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180
 QY 181 RRYSETYGTVTNITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 DB 181 RRYSETYGTVTNITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 QY 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300
 DB 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300
 QY 301 ADAASLETNNRITAVTNPESIVPVLCPGRLQDAXVENPEAGQYMGNIIVTFTPSQT 360
 DB 301 ADAASLETNNRITAVTNPESIVPVLCPGRLQDAXVENPEAGQYMGNIIVTFTPSQT 360
 QY 361 L 361
 DB 361 L 361
 RESULT 2
 ABU09014
 ID ABU09014 standard; protein; 361 AA.
 XX
 CC ABU09014;
 XX
 DT 16-JUN-2003 (first entry)
 XX
 DE Csa operon recombinant product, CsaE.
 XX
 KW CsaE; csa operon; CS4 pilus; bacterial pili protein; ETEC; CS4;
 KW enterotoxigenic Escherichia coli; multivalent Shigella-ETEC; diarrhoea;
 KW Shigella; antibacterial.
 XX
 OS Escherichia coli.
 XX
 EN US2002176968-A1.
 XX
 PD 28-NOV-2002.
 XX
 PF 20-APR-2001; 2001US-00839894.
 XX
 PP 20-APR-2000; 2000US-0198626P.
 XX
 XX (ALTB/) ALTBOUT Z.
 PA (LEVI/) LEVINE M M.
 PA (BARR/) BARRY E M.
 XX
 PI Altboum Z, Levine M, Barry E;
 XX
 DR WPI; 2003-352604/33.

DR N-PSDB; ABX93757.

XX Novel isolated nucleic acid comprising csa operon encoding proteins
 PT required for producing CS4 pili, useful for generating immune response in
 PT vertebrate against the enterotoxigenic Escherichia coli.

XX Claim 10; Page 31-32; 58pp; English.

XX The invention relates to an isolated nucleotide sequence comprising a csa
 CC operon (encoding proteins required for producing CS4 pili) or its
 CC functional fragment. An immunogenic composition comprising a recombinant
 CC product of a csa operon and a carrier, is useful for generating an immune
 CC response in a subject, which involves contacting the subject with the
 CC immunogenic composition. The recombinant product of the composition is
 CC the CS4 antigen (bacterial pili protein) and is provided in an acellular
 CC or cellular composition. The nucleic acid is useful for producing a
 CC polypeptide product from a csa operon or functional fragment, which
 CC involves providing a nucleic acid, introducing the csa operon in an
 CC expression vector, such that a recombinant host cell is produced and
 CC subjecting the recombinant host cell to conditions such that a protein
 CC from the csa operon is expressed. The nucleic acid encoding at least an
 CC immunogenic portion of the csa operon or a polypeptide encoded by the
 CC nucleic acid is useful for generating an immune response in a vertebrate
 CC against ETEC (enterotoxigenic Escherichia coli). The nucleic acid is also
 CC useful for identifying polynucleotides encoding other proteins with
 CC biological functions similar to that of the csa operon and for creating a
 CC multivalent Shigella-ETEC immunogenic composition that will protect from
 CC diarrhoea caused by either Shigella or CS4 expressing ETEC strains. This
 CC sequence represents a recombinant product of a csa operon, CsaE

XX Sequence 361 AA;

Query Match 100.0%; Score 1886; DB 6; Length 361;
 Best Local Similarity 100.0%; Pred. No. 1.6e-168;
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
 DB 1 MNKILFIPTLFFSSVLFTFAVSADKIPGDESITNIFGPRDRNESSPKHNILNNHITAYSE 60
 QY 61 SHLYDRMTFLCLSSHTNLNGACPTSENPSSSSVSGETNITLQFTKRSILKRELQIKGY 120
 DB 61 SHLYDRMTFLCLSSHTNLNGACPTSENPSSSSVSGETNITLQFTKRSILKRELQIKGY 120
 QY 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180
 DB 121 KOLLFKSVNCPGSLTNSAHFNCKNAASGASLYLYIPAGELKNLPFGGIWDATLKLVRK 180
 QY 181 RRYSETYGTVTNITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 DB 181 RRYSETYGTVTNITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGTYIGRNSVDMCFY 240
 QY 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300
 DB 241 DGYSTNSSSLEIRFQDNNPKSDGKFYLRKINDTKEIAYTSLLLAGKSLTPNGTSLNI 300
 QY 301 ADAASLETNNRITAVTNPESIVPVLCPGRLQDAXVENPEAGQYMGNIIVTFTPSQT 360
 DB 301 ADAASLETNNRITAVTNPESIVPVLCPGRLQDAXVENPEAGQYMGNIIVTFTPSQT 360
 QY 361 L 361
 DB 361 L 361

RESULT 3
 ABU09015
 ID ABU09015 standard; protein; 1715 AA.
 XX
 AC ABU09015;
 XX
 DT 16-JUN-2003 (first entry)
 XX

Csa operon related protein.

Csa operon; CS4 pilus; bacterial pili protein; ETEC; CS4;
enterotoxigenic Escherichia coli; multivalent Shigella-ETEC; diarrhoea;
Shigella; antibacterial.

Escherichia coli.

US2002176868-A1.

28-NCV-2002.

20-APR-2001; 2001US-0083989A.

20-APR-2000; 2000US-0198626P.

(ALTE/) ALTBOM Z.

(LEVI/) LEVINE M.M.

(BARR/) BARRY E.M.

Altbaum Z, Levine MM, Barry EM;

WPI; 2003-352604/33.

Novel isolated nucleic acid comprising csa operon encoding proteins
required for producing CS4 pili, useful for generating immune response in
vertebrate against the enterotoxigenic Escherichia coli.

Disclosure; Page 35-41; 58pp; English.

The invention relates to an isolated nucleotide sequence comprising a csa
operon (encoding proteins required for producing CS4 pili) or its
functional fragment. An immunogenic composition comprising a recombinant
product of a csa operon and a carrier, is useful for generating an immune
response in a subject, which involves contacting the subject with the
immunogenic composition. The recombinant product of the composition is
the CS4 antigen (bacterial pili protein) and is provided in an acellular
or cellular composition. The nucleic acid is useful for producing a
polypeptide product from a csa operon or functional fragment, which
involves providing a nucleic acid, introducing the csa operon in an
expression vector, such that a recombinant host cell is produced and
subjecting the recombinant host cell to conditions such that a protein
from the csa operon is expressed. The nucleic acid encoding at least an
immunogenic portion of the csa operon or a polypeptide encoded by the
nucleic acid is useful for generating an immune response in a vertebrate
against ETEC (enterotoxigenic Escherichia coli). The nucleic acid is also
useful for identifying polynucleotides encoding other proteins with
biological functions similar to that of the csa operon and for creating a
multivalent Shigella-ETEC immunogenic composition that will protect from
diarrhoea caused by either Shigella or CS4 expressing ETEC strains. This
sequence represents a csa operon related protein of the invention

Sequence 1715 AA;

Query Match 99.7%; Score 1881; DB 6; Length 1715;

Best Local Similarity 100.0%; Pred. No. 4.7e-167;

Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 NKILFIPTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESPKNILNHHITAYSS 61

1264 NKILFIPTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESPKNILNHHITAYSS 1323

62 HTLYDRMTFLCLSSHTNLGACPTSPNPSSSVSGETNITLQTEKSLIKRELQIKGYK 121

1324 HTLYDRMTFLCLSSHTNLGACPTSPNPSSSVSGETNITLQTEKSLIKRELQIKGYK 1383

122 QILFKVNCPSGUTLNSAHFNCNKAASGASLYIYPAGELKXLPFGGIWDATLKLVRKR 181

1384 QILFKVNCPSGUTLNSAHFNCNKAASGASLYIYPAGELKXLPFGGIWDATLKLVRKR 1443

182 RYSETYGTYTINITIKLTKGNIQIWLPOFKSDARVDLNLRTGGGTYIGRNSVDXCFYD 241

1444 RYSETYGTYTINITIKLTKGNIQIWLPOFKSDARVDLNLRTGGGTYIGRNSVDXCFYD 1503

QY 242 GYSTNSSLEIRFQDNPKSDKFKYLRKINDTKETIAVYTLISLLLAGKSLTPNGTSLNIA 301
DB 1504 GYSTNSSLEIRFQDNPKSDKFKYLRKINDTKETIAVYTLISLLLAGKSLTPNGTSLNIA 1563
QY 303 DAASLETNNRITAVTMEISVPVLCWPEGRQLDAKVENPEAGQTMGNINVTFTPSQTL 361
DB 1564 DAASLETNNRITAVTMEISVPVLCWPEGRQLDAKVENPEAGQTMGNINVTFTPSQTL 1623

RESULT 4

AAV22326

ID AAY22326 standard; protein; 364 AA.

XX AAY22326;

AC AAY22326;

DT 22-SEP-1999 (first entry)

DE Pili protein CotD.

XX CS2 gene cluster; CotA; CotB; CotC; CotD; pili protein; immunogen;

KW enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;

KW enteric infection; therapy.

XX Escherichia coli.

XX US5932715-A.

XX 03-AUG-1999.

XX 07-JUN-1995; 95US-00483101.

XX 07-JUN-1995; 95US-00483101.

XX (UYEM-) UNTV EMORY.

XX Caron J. Froehlich B, Scott JR;

XX WFI; 1999-443623/37.

XX N-PSDB; AAX84848.

XX Isolated nucleic acids encoding Escherichia coli CS2 pili proteins

XX useful for vaccinating against diarrheal diseases caused by Escherichia

XX coli.

XX Claim 3; Col 45-48; 35pp; English.

XX This sequence represents a CS2 pili protein of the invention, encoded by

XX (which also encodes CotA, CotC, and CotB). CS2 pili are long

XX proteinaceous molecules thought to mediate attachment of enterotoxigenic

XX E. coli (ETEC) to and/or promote colonisation of the human upper

XX intestine. The CS2 gene cluster may be used to produce immunogens for

XX vaccinating patients against diarrhoeal diseases caused by ETEC bacteria.

XX This type of enteric infection is a major cause of death among infants in

XX developing countries and in immunocompromised (e.g. Acquired Immune

XX Deficiency Syndrome (AIDS)) or elderly adults. The vaccine comprises more

XX than 1 antigenic determinant (epitopes) from more than 1 pili type to be

XX effective against more than 1 type of ETEC infection

XX Sequence 364 AA;

Query Match 49.2%; Score 927.5; DB 2; Length 364;

Best Local Similarity 48.1%; Pred. No. 2.4e-78;

Matches 176; Conservative 65; Mismatches 118; Indels 7; Gaps 3;

QY 1 MKKILFIPTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRN--ESSPKHNLNHHITAY 58

DB 1 LKKVIFVLSMF--CSQVYGQSWHINVEAGSINKTESICPIDR3AAASYPAFYFHEHVAGY 60

QY 59 SESHLYDRMTFLCLSSHTNLGACPTSPNPSSSVSGETNITLQTEKSLIKRELQIK 118

DB 61 NKCHSLFDRMTFLCMSSTDAKGACPTGNSKES--QGETNKLIFTEKSKSLARKTLNKL 118

CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 GC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1335 AA;

Query Match 6.4%; Score 120; DB 6; Length 1335;
 Best Local Similarity 20.4%; Pred. No. 0.15;
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

QY 13 SSVLTFEFAVADKIPGDESITNIFGPRDRNESSPKHNILNHTAY---SESHLYDRMT 69
 DB 746 SAKIATLSASNNVLANEAMNTVSVNVADEGS---NPINDHTVFAVLGSATSFNNON 802
 QY 70 FICLSHNTLNGACPTSENPFSSSVSGETNITLQFTEKRSLI-----K 112
 DB 803 ---HAKTDVNGLA-TFDLKSQKQEDNTVEVLNGVKQTLIVSFVGDSSTAQVDLQSK 857
 QY 113 RELQIKGY-KQLLFKSVNCPGLTNSAHFNCKNKAASGASLYIYPAGELKNLPFGGIW 171
 DB 858 NSVADGNSATMTATVRDAKGNLLNDVKTFFNVNSAAKLSQTEVNSHD-----GTA 910
 QY 172 DATLKLVRKRYSETGYTITNITIKLTPKGNIQIWLPOFKSDARVDLNLRTGGTYIG 231
 DB 911 TAYLT-----SLKNGDYTVASVSGSQNQVIFIGDQTAALTLSV-PSGDTV-- 960
 QY 232 RNSVDMCFYDGYSTNSSLE---IRFQNNPKSGCKFYLRKINDTKETIAYTSLILAG 287
 DB 961 -----TNTAPLHMTATLQCKNGNPLKDKKEITFSVPND-----VASRFSISNSG 1003
 QY 286 KSLTFNTHSL-----NIADAASLETWNRTITATMTEIPEISVPVL 327
 DB 1004 KGMTDSNGTAIASLCTAGTHMITARLANSNVSDTPQMTFFVADKRAVAVVLQTSKAEII 1063
 QY 328 WFG--RLQDAAKVENPEAGQYGNINVTFS 357
 DB 1064 GNGVDETTLTAVKDP-FDNVKNLSVVFATS 1094

RESULT 7

AY50814
 ID AY50814 standard; protein; 770 AA.

XX AY50814;

XX AY50814;

DT 17-FEB-2000 (first entry)

XX

DE D. melanogaster acetyl-cho-line receptor protein from clone Da7.
 XX
 KW Acetyl-cho-line receptor; nicotinic; insect; insecticide; screening;
 KW neurotransmission; plant protection agent; conductance; Achr.
 XX
 OS Drosophila melanogaster.
 XX
 PN DE19819829-A1.
 XX
 PD 11-NOV-1999.
 XX
 PD 04-MAY-1998; 98DE-010-9829.
 XX
 PR 04-MAY-1998; 98DE-010-9829.
 XX
 PA (FAR) BAYER AG.
 XX
 PI Adamczewski M, Oellers N, Schulte T;
 XX
 XX WPI: 2000-014207/02.
 DR N-PSDB; AA224475.
 XX

New nucleic acid encoding a nicotinic acetylcholine receptor from
 insects, used to identify potential insecticides.

Example 1a; Page 12-14; 26pp; German.

This invention describes a novel nucleic acid (NA) encoding a nicotinic
 acetyl-cho-line receptor (I) from insects which can be used as an
 insecticide. Inhibitors of (I) interfere with neurotransmission. (I)
 (also vectors containing it, its regulatory regions, and antibodies
 directed against (I)-encoded proteins) are used to screen for: (a) plant
 protection agents that alter conductance of ACHR, potentially useful as
 insecticides, or (b) genes which encode polypeptides that are involved in
 formation of functionally related ACHR in insects. (I) are also used to
 isolate and characterize the specified regulatory regions and for
 recombinant production of (II). This sequence represents an acetyl-
 choline receptor isolated from *Drosophila melanogaster*

SQ Sequence 770 AA;

Query Match 6.3%; Score 119.5; DB 3; Length 770;
 Best Local Similarity 20.4%; Pred. No. 0.074;
 Matches 33; Conservative 48; Mismatches 140; Indels 135; Gaps 20;

QY 9 ILFPSSVLETFVAADKIPGDESITNIFGPRDRNESSPKHNILNNE----- 54
 DB 243 TIATFISYLGSAQAQKNSSSSSSSN---SSNNSSTQILNGLNKESWIFLLIYLNLISAK 304
 QY 55 --ITAVSBSHTLYDRMTFLCSSHNTLNGACPTSENPPSSSVSGETNITLQFTEKRSLIK 112
 DB 305 VCIAGYHEKLLHD-----LLDPYTLERFVLNEDPQLSFGTLMOIYDVERKNQLIV 359
 QY 113 RELQIKGYKQLLFKSVNCPGSLTNSAHFNCKNKAASGASLYIYPAGELKNLPFGGIWD 172
 DB 360 TNVWLK-----LEWDMNLRWNTSDYGVVK-----DLR-IPRHRIWK 395
 QY 173 ATLKLVRKRYSET---GYTYTINITIKLTPKGNIQIWLPO--FKSDARVDLNLRTGG 226
 DB 395 PDVLM-----YNSADEGFDGTQTNVVR--NNGSC-LYVPEGIFKSCIKDITWFP--- 444
 QY 227 GTYIGKNSVDMCF---YDGYSTNSSSLEIRFQDNNPKS-----DGKPYLRKINDTKTE 276
 DB 445 ---FDDQRCMCFGSGWTYDGF-----QLDLQDDETDGDISSYVINGEWELLGVGKRN 496
 QY 277 IAYT-----SLILAGKSLTPTNGTSLNIADA 303
 DB 497 IYVNCPEPYIDITAIIRRTLYFFNLIIPCVLIASVALLGFTLPDSCSEKLSGVT 556
 QY 304 ASLE-TNNWRITATMPEIS--VPV-----ICW-----PGR 331
 DB 557 ILLSLTVFLNVAETMPATSDAIVLIRIVFLCWLPLWILRMRPGR 602

```
RESULT 8
ABB82573
ID ABB82573 standard; protein; 1386 AA.
XX
XX ABB82573;
AC
XX
XX 04-FEB-2003 (first entry)
DT
XX
XX H. influenzae BASB223 or BASB224 polypeptide sequence.
DE
XX
XX BASB223, BASB224; high molecular weight protein A; immunostimulant;
KW antibacterial; vaccine; medicament; bacterium.
XX
XX Haemophilus influenzae.
OS
XX
XX WO200279237-A2.
PN
XX
XX 10-OCT-2002.
PD
XX
XX 12-MAR-2002; 2002WO-EP003210.
PF
XX
XX 13-MAR-2001; 2001GB-00006155.
PR
XX
XX 13-MAR-2001; 2001GB-00006155.
XX
XX (GLAX ) GLAXOSMITHKLINE BIOLOGICALS SA.
PA
XX
XX Thonard J;
PI
XX
XX WPI; 2003-040650/03.
DR
XX
XX N-PSDB; ABV74997.
XX
XX New BASB223 and BASB224 polypeptides and polynucleotides, useful for
PT preparing a medicament for generating an immune response in an animal, or
PT as research reagents and materials for discovering treatments for human
PT diseases.
XX
XX Claim 3; Page 71; 109pp; English.
PS
XX
XX The invention relates to novel polypeptides BASB223 and BASB224, isolated
CC from non-typeable H. influenzae (NTHi). The polypeptides are related to
CC H. influenzae high molecular weight protein A (HmwA). Compositions
CC comprising the BASB223 and BASB224 polypeptides and polynucleotides are
CC useful for preparing a medicament for generating an immune response in an
CC animal. The BASB223 and BASB224 polypeptides and polynucleotides may be
CC used as research reagents and materials for discovery of treatments and
CC diagnosis for diseases, particularly human diseases, and as immunogens
CC for producing antibodies immunospecific the such polypeptides and
CC polynucleotides. The polynucleotides may be used in the discovery and
CC development of antibacterial compounds, in genetic immunization, in the
CC development of prophylactic or therapeutic agents for bacterial
CC infection, particularly non-typeable H. influenzae. The present sequence
CC represents the BASB223 or BASB224 polypeptide partial sequence
XX
XX Sequence 1386 AA;
SQ
Query Match 6.2%; Score 117.5; DB 6; Length 1386;
Best Local Similarity 20.4%; Pred. No. 0.27;
Matches 81; Conservative 62; Mismatches 136; Indels 119; Gaps 17;
QY 45 SPANILNN---HITAYSESHLYYDRMTPLCLSSHTLNGACPTSENPPSSSVSGFT---- 98
DB 697 SSRINVTNGSSLSITGDMPAKIFDKNDLV-NA-----TNGNVSITEVEGSDTKL 747
QY 99 -----NITLQFTKRLIKRELQIKGYKQLLPKSVNCPGLTINSAHNC 143
DB 748 EYGLVADGNITVEGGNVLGSKNAKTHITANSVK-----NANLILSSANFV 796
QY 144 NKNAAS-GASLXLY-PAGELKNLPFGIW-----DNLKLRVRRRYSETYGTYT 191
DB 797 HKGALTIGGS-----ANIQGNITANGDTVEAGDVIVSDDAKFAETKNNLIT-GTFT 849
QY 192 IN-----ITIK-----LTDKNGTIQW-----LPQKSDA 215
```

```
850 NNGTSEINIKQGVVKLOQDITNNNGNLNITTNASVNOXTIINGNIINKKGDLINIKDKANA 909
216 RVDL--NLRPTGGGTIGRNSVDVDMCFYDGYSTNGSSLEIRFODNNPKSD-GKFYLRKIND 272
910 BIQGGNISQKEGNTLISSDKINI-----TKRIEIKADTQGNSDSGVANSANLITI 960
273 DTKEIAYTLSLLLAG---KSLTPNGTSLNIADAAASLETNNRITAVTMTPEISVPVLCWP 329
931 KTKELTLDNLNISGFNKABITAKDNSDLIIGKASSDNSNAKQITFDKVKDKIS--AGN 1018
330 GRLOLDKAVENPFA-----QQYMGNIINVTTPSSQIL 361
1019 HNVTLNSKVTSNSDGTGNGSDNNIGLTISAKDVTV 1056

RESULT 9
ABB82570
ID ABB82570 standard; protein; 1449 AA.
XX
XX ABB82570;
AC
XX
XX 04-FEB-2003 (first entry)
DT
XX
XX H. influenzae BASB223 polypeptide.
DE
XX
XX BASB223; BASB224; high molecular weight protein A; immunostimulant;
KW antibacterial; vaccine; medicament; bacterium.
XX
XX Haemophilus influenzae.
OS
XX
XX WO200279237-A2.
PN
XX
XX 10-OCT-2002.
PD
XX
XX 12-MAR-2002; 2002WO-EP003210.
PF
XX
XX 13-MAR-2001; 2001GB-00006155.
PR
XX
XX 13-MAR-2001; 2001GB-00006155.
XX
XX (GLAX ) GLAXOSMITHKLINE BIOLOGICALS SA.
PA
XX
XX Thonard J;
PI
XX
XX WPI; 2003-040650/03.
DR
XX
XX N-PSDE; ABV74994.
XX
XX New BASB223 and BASB224 polypeptides and polynucleotides, useful for
PT preparing a medicament for generating an immune response in an animal, or
PT as research reagents and materials for discovering treatments for human
PT diseases.
XX
XX Claim 3; Page 67-68; 109pp; English.
PS
XX
XX The invention relates to novel polypeptides BASB223 and BASB224, isolated
CC from non-typeable H. influenzae (NTHi). The polypeptides are related to
CC H. influenzae high molecular weight protein A (HmwA). Compositions
CC comprising the BASB223 and BASB224 polypeptides and polynucleotides are
CC useful for preparing a medicament for generating an immune response in an
CC animal. The BASB223 and BASB224 polypeptides and polynucleotides may be
CC used as research reagents and materials for discovery of treatments and
CC diagnosis for diseases, particularly human diseases, and as immunogens
CC for producing antibodies immunospecific the such polypeptides and
CC polynucleotides. The polynucleotides may be used in the discovery and
CC development of antibacterial compounds, in genetic immunization, in the
CC development of prophylactic or therapeutic agents for bacterial
CC infection, particularly non-typeable H. influenzae. The present sequence
CC represents the BASB223 polypeptide
XX
XX Sequence 1449 AA;
SQ
Query Match 6.2%; Score 117.5; DB 6; Length 1449;
Best Local Similarity 20.4%; Pred. No. 0.29;
```

Matches 81; Conservative 62; Mismatches 136; Indels 119; Gaps 17;

QY 45 SPKHLNKK---HITAYSESHLYDRMTFLCLSSHTNLNGACPTSPSSSVSGPT--- 98

Db 760 SSRINVTGSS--SITGDMPAKKIFDIKNDLVINA-----TNSNVSITEVEGDTKL 810

QY 99 -----NITLOFTEKRSIKRELQIKGYKQLLPKSVNCPGLTNSAHENC 143

Db 811 EYGLVADGNITVEGNTVLGSKAKTHITKNSVK-----NANLTSSANFNV 859

QY 144 NKAAAS-GASLYLYIPAGELKNLPFGGIW-----DATLKLVRKRYSETYGT 191

Db 860 HKGALTIGGS-----ANIQGNLTAKGDTVEVAGDVIVSDDAKFAETKNLNIT--GT 912

QY 192 IN-----ITIK-----LTKGNITQIW-----LPOPKSDA 215

Db 913 NNGTSEINIKQGVVKLOQGITNNGN--NITNASVNOKTINGNITNKKGLNLIKDIKANA 972

QY 216 RVDL--NLKPTGGCTVIGRNSVDMCFVDGYSTNSSSLEIRFQDNNPKSP--GKPYLRKIND 272

Db 973 ETQIGNISQKEGNTLISDKINI-----TKRIEKADTDQNSDGSVGSANANLTI 1023

QY 273 DKYEAYTUSLLAG---KSLPTNGTSLNIADAASLETNMRITAVTPEISVPVLCWP 329

Db 1024 KTKELTLTDKLNISGPNKAEITAKDNSDLIIGKASSDNSNAKQITFDKVKDSKIS--ACN 1081

QY 330 GLQLDAKVNDEA-----QVMGNINVTFTPSSOTL 361

Db 1082 HNVTLNSKVTESDGSSTGSGSDNNGTISAKDVTV 1119

RESULT 10

AA47575

ID AAR47575 standard; protein; 650 AA.

XX AAR47575;

AC AAR47575;

DT 25-MAR-2003 (revised)

DT 19-JUL-1994 (first entry)

XX Alpha-agglutinin of Saccharomyces cerevisiae.

DE Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;

KW Major cell wall protein; glycosyl-phosphatidyl-inositol;

KW anchoring protein; alpha factor; alpha-agglutinin; invertase; inulinase;

KW alpha-amylase; Saccharomyces cerevisiae; enzymatic process; fermentation;

KW biodegradation; catalysis.

CS Saccharomyces cerevisiae.

XX WO9401567-A1.

PN 20-JAN-1994.

PD 07-JUL-1993; 93WO-EP001763.

XX 08-JUL-1992; 92EP-00202080.

PR 14-DEC-1992; 92EP-00203899.

XX (UNL) UNILEVER PLC.

PA (UNL) UNILEVER NV.

PI Klis FM, Schreuder MP, Toschka H, Verrips CT;

XX WPI; 1994-035071/04.

DR N-PSDB; AAQ54012.

XX Immobilisation of enzymes to microbial cell wall - by prodn. of fusion

PT protein of enzyme linked to anchoring protein.

XX Example 2; Page 32-39; 99pp; English.

PS The alpha-agglutinin is used in a method to immobilise enzymes to a

CC

microbial cell wall. The coding sequence is used in the production of a recombinant polynucleotide which comprises a structural gene encoding a protein with catalytic activity and at least part of a gene encoding at least the C-terminus of a protein capable of anchoring in a eukaryotic or prokaryotic cell wall. The anchoring fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a proteinase of lactic acid bacteria. The recombinant polynucleotide preferably also comprises a sequence encoding a signal peptide to ensure secretion of the expressed product. The signal peptide is preferably derived from glycosyl-phosphatidyl-inositol, anchoring protein, alpha factor, alpha-agglutinin, invertase or inulinase, alpha-amylase of Bacillus or proteinases of lactic acid bacteria. The host microorganism can be used for performing enzymatic processes on an industrial scale. (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 650 AA;

Query Match 6.0%; Score 114; DB 2; Length 650;

Best Local Similarity 20.9%; Pred. No. 0.19;

Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFIETLFSSVL-----VSADKIPG-----DESTNIFGPRDRNE--- 43

Db 3 ILWLFLALASAININDITFSNLBITPLTANKPQDQGTATDFSLADASSIREGDEFTL 67

QY 44 SSP---KNILNNHITA-----YSESHLYDRMTFLC-----LSSHNTLN 80

Db 63 SMPHVYRIKLLNSSQTATISLADGTEAFKCVYVQQAALYENTTCTAQNLDSSVNTID 127

QY 81 GACPTSENPSSSVSGEINTILOFTEKRSIKRELQIKGYKQLLPKSVNCPGLTNSA 139

Db 123 GSITFELNPFSGGSSYEVEL-----ENAKFFKSGGPMVLKLGQMSDVNVNDFEAAFTENVF 182

QY 143 HENCNKAASGASLYL--YIPAGELKNLPFGGIWDATLKLVRKRYSETYGTYNITIK 197

Db 183 HGRSTGYGSFESYHLGYCPNGYF---LGG---TEKI-----DYSSNNNVDL- 225

QY 193 LTDKNIQI-----WLPQKSDARVDNLNRPFGGTIGRN---SVMCFYDG--- 242

Db 225 --DCSSVQVYSSNDFNWFQSYNDTNADV-----TCFGSNLWITLDBKLYDGEML 275

QY 243 -----YSTNSSLE--IRFQDNN--PKSDGKF 265

Db 275 WYNALQSLPANVNTIDHALEFOYTCLDTIANTTYATQFTTREFIVYQGRNLGTASAKSS 335

QY 265 YLRKINDTKBI---AYTILS-----ILLAKSLTPTNGTSLNIADAA 304

Db 335 FISTVTTDLTSLNTSAYSTGISTVETGNRTTSEVISHVVTSTKLSPTATSLTIAQTS 395

QY 305 SLETNWN-----RIVAVNMPESVPLVLCWPGRL 332

Db 396 IYSTDSNITVGTDIHTTSEVISDVETISRETASTVVAAPTSTGTGTGAM 444

RESULT 11

ABU33527

ID ABU33527 standard; protein; 513 AA.

XX ABU33527;

AC ABU33527;

DT 19-JUN-2003 (first entry)

XX Protein encoded by prokaryotic essential gene #19054.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

KW Legionella pneumophila.

XX WO20027183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US09107.

PF

proteins of Haemophilus which are used as a vaccine to provide protection against Haemophilus induced diseases in humans.

Claim 12; Fig 21A-O; 307pp; English.

The invention relates to the recombinant production of Haemophilus influenzae high molecular weight (HMW) proteins in Escherichia coli. The expression construct used to effect recombinant expression comprises a promoter functional in E. coli (e.g., the T7 promoter) operably linked to a modified hmwaABC operon from a non-typeable (non-encapsulated) H. influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene clusters termed hmwa1ABC and hmwa2ABC. Each hmwaABC operon comprises hmwaA, hmwaB and hmwaC genes. The hmwaA genes encode the structural HMWA proteins and the hmwaB and hmwaC genes encode accessory proteins which are responsible for post-translational processing and secretion of the HMWA proteins. The modified hmwaABC operon used in the expression construct of the invention contains an A gene modified such that it encodes only the mature HMWA. The invention also discloses hmwa genes (AAAS2175-AS2198) and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae strains Jovc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and vectors are used for the production of recombinant H. influenzae HMW proteins which can be used as vaccines to mediate a humoral or cell-mediated immune response to provide protection against diseases in humans caused by H. influenzae (e.g., otitis media, epiglottitis, pneumonia and tracheobronchitis). The HMW proteins are also useful as antigens in immunoassays for detecting antibodies against Haemophilus, HMW proteins and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence represents an HMWA protein from a non-typeable strain of H. influenzae. (Updated on 12-SEP-2003 to standardise OS field)

Sequence 1011 AA;

Query Match 5.9%; Score 111.5; DB 3; Length 1011;
Best Local Similarity 20.3%; Pred. No. 0.63;
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

QY 60 ESHLYDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQFTKRS-LIKRELQIK 118
DB 369 EKNAIFSTHNLITLGGNVITLGGNSNNKGNININASKANVTLOAHAGTSHLDKE--- 424
QY 119 GYKOLLFKSVKCPGLTL--NSAHFNCNKAASGASLYLYIPAGEL-KNLPFGGIW--DA 173
DB 425 --RTLTLGNVSVGGNLIIGSNNAHIDGNLS-ABSAKP-----QCKTNNNLNITGFTTNG 477
QY 174 TLKLRVRRYSEYTYGTYTINITIKLTD-----KGNI-----QIWLFPQKSDARVD 218
DB 478 TADINIKQGVVKLQGDITNNGNINITTNASVNOKTIINGNITNKKGLNLIKDIKANAEIQ 537
QY 219 L--NLRPTGGTIGRNSVDMCFYDGYSTNSSLSEIRFQDNNPKSD-GKFLRKINDDTK 275
DB 538 IGGNISQKGNLTISSDKINI-----TKRIEKADTDQGNSDSGVASNANLTITKT 588
QY 276 EIAYTSLLLAG---KSLTPTNGTSLNIADAASLETWNRSITAVTMEIISVPVLCWQRL 332
DB 539 ELTLDNLNISGFNKAEBITAKDNDLLIGKASSDNSNAKQITFDKVKDSKIS--AGHNIV 646
QY 333 QLDKAVENPEA-----GOYMGNIINVTFTPSQTL 361
DB 647 TLNSKVETSDSGTNGSDNNIGLTISAKDVT 691

RESULT 14
AAB01832
ID AAB01832 standard; protein; 1529 AA.
XX AAR41732;
AC AAR41732;
XX 25-MAR-2003 (revised)
DT 26-APR-1994 (first entry)
XX High molecular weight protein 4 (HMW4).
DE

CC and/or HMW peptides. The nucleotide sequences encoding the HMW proteins can be used to isolate and clone hmw genes from other non-typeable strains of Haemophilus via hybridisation reactions. The present sequence CC represents a mature HMWA protein from a non-typeable strain of H. influenzae. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ Sequence 1005 AA;

Query Match 5.9%; Score 111.5; DB 3; Length 1005;
Best Local Similarity 20.3%; Pred. No. 0.62;
Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

QY 60 ESHLYDRMTFLCLSSHNTLNGACPTSENPSSSVSGETNITLQFTKRS-LIKRELQIK 118
DB 363 EKNAIFSTHNLITLGGNVITLGGNSNNKGNININASKANVTLOAHAGTSHLDKE--- 418
QY 119 GYKOLLFKSVKCPGLTL--NSAHFNCNKAASGASLYLYIPAGEL-KNLPFGGIW--DA 173
DB 419 --RTLTLGNVSVGGNLIIGSNNAHIDGNLS-ABSAKP-----QCKTNNNLNITGFTTNG 471
QY 174 TLKLRVRRYSEYTYGTYTINITIKLTD-----KGNI-----QIWLFPQKSDARVD 218
DB 472 TADINIKQGVVKLQGDITNNGNINITTNASVNOKTIINGNITNKKGLNLIKDIKANAEIQ 531
QY 219 L--NLRPTGGTIGRNSVDMCFYDGYSTNSSLSEIRFQDNNPKSD-GKFLRKINDDTK 275
DB 532 IGGNISQKGNLTISSDKINI-----TKRIEKADTDQGNSDSGVASNANLTITKT 582
QY 276 EIAYTSLLLAG---KSLTPTNGTSLNIADAASLETWNRSITAVTMEIISVPVLCWQRL 332
DB 583 ELTLDNLNISGFNKAEBITAKDNDLLIGKASSDNSNAKQITFDKVKDSKIS--AGHNIV 640
QY 333 QLDKAVENPEA-----GOYMGNIINVTFTPSQTL 361
DB 641 TLNSKVETSDSGTNGSDNNIGLTISAKDVT 675

RESULT 13
AAB01832
ID AAB01832 standard; protein; 1011 AA.
XX AAB01832;
AC AAB01832;
DT 12-SEP-2003 (revised)
DT 11-SEP-2000 (first entry)
XX Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.
XX HMW protein; hmwa1; hmwa2; high molecular weight;
KW non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW recombinant production; Escherichia coli; antibacterial; vaccine;
KW human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW detection; diagnosis.
XX Haemophilus influenzae; strain K21.
XX WO200020609-A2.
XX 13-APR-2000.
XX 07-OCT-1999; 99WO-CA000938.
XX 07-OCT-1998; 98US-00167568.
XX 08-DEC-1998; 98US-00206942.
XX (CONN-) CONNAUGET LAB LTD.
XX Loosmore SM, Yang Y, Klein ME;
XX WPI; 2000-303789/26.
XX N-PSDB; AAA52181.
XX Nucleic acid molecule for producing recombinant high molecular weight
PT

XX HMW; high molecular weight protein; virus; vaccine; influenza; epitope;
 KW immunity; haemophilus influenzae.
 XX
 XX haemophilus influenzae.
 OS
 XX WO9319090-A1.
 PN
 XX 30-SEP-1993.
 PD
 XX 16-MAR-1993; 93WO-US002166.
 XX
 XX 16-MAR-1992; 92GB-00005704.
 PR
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA
 XX Barenkamp SJ;
 PI
 XX WPI; 1993-320683/40.
 DR
 XX N-PSDB; AAQ49511.
 DR
 XX High molecular weight surface proteins - of non-typeable haemophilus
 PT which exhibit immunogenic properties.
 FT
 XX Claim 6; Fig 10; 100pp; English.
 PS
 XX The isolation and purification of the high molecular weight protein
 CC enables the identification of the major protective epitopes of the
 CC protein by conventional epitope mapping. These epitopes can then be
 CC synthesised using standard techniques and incorporated into fully
 CC synthetic or recombinant vaccines. (Updated on 25-MAR-2003 to correct FN
 CC field.)
 CC
 XX Sequence 1529 AA;
 SQ
 Query Match 5.9%; Score 110.5; DB 2; Length 1529;
 Best Local Similarity 20.7%; Pred. No. 1.4;
 Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;
 QY 48 HNILNKHITAYSESHTLYDRMTFLCLSSHTNLGACPTSENPS--SVSGETNITL 102
 Db 744 YNEYSKH-AINSSHNJ-----TILGNNVTLGG-----ENSSSITGNINITKANVIL 790
 QY 103 QF-----TEKRSLIKRELOIKGYKQLLPKSVNCPGLTNSAHFNCNKAASGLY 154
 Db 791 QADTSNSNTGLKRTLTGLNISVEGNLSLTGANANIVGNLSI--ABDSTFKGEAS----- 843
 QY 155 LYIPAGELKNLPFGGLW--DAILKLRVKRRYSYTYGT--INITIKLT----- 199
 Db 844 -----DNLNITGTFTNNGTANINIKQGVKLGDIINKGGLNITNASGTQKTIING 895
 QY 200 -----DKGNIQIWLPOPKSDARVDL--NLRP TGGGTVIGRNSVDYCFYDGYSTKSSLEIR 253
 Db 896 NITNEXGLDLNI--KNIKADAEIQIGNISQKEGNLTSSDKVNI-----TNQITIKAG 946
 QY 254 FQ-----DNNPKSDGKFLRKINDDTKEIATYLSLLAG---KSLTPNFTGSLNIADAASL 306
 Db 947 VEGCRSDSSEANANLTIO-----TKELKLAGDLNISGFNKAEITAKGSDLTIGNASGG 1001
 QY 307 ETNKNRITAVTMEPISVPLVCWPG--RLQLDKAYENPEAGQYMGINIVT 353
 Db 1002 NADAKK---VTFDKVKDSKISTGHNVTLSNVEKTSNGSSNAGNDNST 1046
 RESULT 15
 AAW30292
 ID AAW30292 standard; protein; 1601 AA.
 XX
 AC AAW30292;
 XX
 XX 17-OCT-2003 (revised)
 DT 14-APR-1998 (first entry)
 XX

DE Non-typeable Haemophilus high mol.wt. surface protein HMW4.
 XX
 KW Non-typeable Haemophilus; high molecular weight surface protein; HMW4;
 KW immunogen; vaccine; otitis media.
 XX
 OS haemophilus influenzae; strain 5.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 372 /note= "encoded by TCT"
 FT Misc-difference 400 /note= "encoded by AAT"
 FT
 XX WO9736914-A1.
 PN
 XX 09-OCT-1997.
 PD
 XX 01-APR-1997; 97WO-US004707.
 PF
 XX 01-APR-1996; 96US-00617697.
 PR
 XX (BARE/) BARENKAMP S J.
 PA
 XX Barenkamp SJ;
 PI
 XX WPI; 1997-503038/46.
 DR
 XX N-PSD3; AAT90993.
 DR
 XX High molecular weight prcteins of non-typeable Haemophilus influenzae -
 PT useful for vaccine production.
 FT
 XX Claim 1; Page 97-102; 183pp; English.
 PS
 XX This protein comprises the high molecular weight surface protein HMW4
 CC (123 kDa) of non-typeable Haemophilus influenzae strain 5 that has the
 CC immunological ability to protect against disease caused by a non-typeable
 CC Haemophilus strain and is characterised by at least one surface-exposed B
 CC -cell epitope that is recognised by monoclonal antibody Ab6. The HMW4
 CC amino acid sequence was deduced from an isolated hmw4 gene (see
 CC AAT90993). HMW1 (see AAW30293), HMW2 (see AAW30294) and HMW3 (see
 CC AAW30291) have also been identified. A conjugate comprising HMW4 linked
 CC to an antigen, hapten or polysaccharide, and a synthetic peptide of 6-150
 CC amino acids corresponding to at least protective epitope of HMW4 are also
 CC claimed. HMW proteins, conjugates and peptides can be used in vaccines,
 CC as immunogens for preparation of antibodies and as antigens for detection
 CC of these antibodies. (Updated on 17-OCT-2003 to standardise OS field)
 CC
 XX Sequence 1601 AA;
 SQ
 Query Match 5.9%; Score 110.5; DB 2; Length 1601;
 Best Local Similarity 20.7%; Pred. No. 1.5;
 Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;
 QY 48 HNILNKHITAYSESHTLYDRMTFLCLSSHTNLGACPTSENPS--SVSGETNITL 102
 Db 816 YNEYSKH-AINSSHNJ-----TILGNNVTLGG-----ENSSSITGNINITKANVIL 862
 QY 103 QF-----TEKRSLIKRELOIKGYKQLLPKSVNCPGLTNSAHFNCNKAASGLY 154
 Db 863 QADTSNSNTGLKRTLTGLNISVEGNLSLTGANANIVGNLSI--ABDSTFKGEAS----- 915
 QY 155 LYIPAGELKNLPFGGLW--DAILKLRVKRRYSYTYGT--INITIKLT----- 199
 Db 916 -----DNLNITGTFTNNGTANINIKQGVKLGDIINKGGLNITNASGTQKTIING 967
 QY 200 -----DKGNIQIWLPOPKSDARVDL--NLRP TGGGTVIGRNSVDYCFYDGYSTKSSLEIR 253
 Db 968 NITNEXGLDLNI--KNIKADAEIQIGNISQKEGNLTSSDKVNI-----TNQITIKAG 1018
 QY 254 FQ-----DNNPKSDGKFLRKINDDTKEIATYLSLLAG---KSLTPNFTGSLNIADAASL 306
 Db 1019 VEGCRSDSSEANANLTIO-----TKELKLAGDLNISGFNKAEITAKGSDLTIGNASGG 1073

Oy 307 ETNNRITATYEBISVPVLCWPG-RLQLDAXVENPEAGQYMGNIINV 353
Db 1074 NADAKK---VTFDKVDSKISTDGHNVTLINSEVKTNGSSNAGNDNST 1118

Search completed: May 6, 2004, 10:17:42
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 10:15:31 ; Search time 20 Seconds
(without alignments)
1736.257 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MNKILFIFLFFSVLTFPA.....EAGQYMNINVTTPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76.*

1: P.r1.*

2: P.r2.*

3: P.r3.*

4: P.r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1558.5	82.6	360	2	D56617	cfab protein precu
2	949.5	50.3	363	2	S49539	CotD protein precu
3	930.5	49.3	364	2	S57937	CotD protein precu
4	275.5	14.6	359	2	AB0541	probable fimbrial
5	120	6.4	1335	2	G90975	probable factor [i
6	120	6.4	2660	2	E85822	probable invasin Z
7	114	6.0	650	2	S28835	alpha-agglutinin -
8	109	5.8	2383	2	D64962	probable membrane
9	108	5.7	524	2	S55097	probable membrane
10	108	5.7	2529	2	B64635	toxin-like outer m
11	107	5.7	304	2	A64904	probable fimbrial
12	107	5.7	304	2	C90892	probable adhesin [
13	107	5.7	304	2	F85725	probable adhesin,
14	106	5.6	671	2	A38109	autolysin - Enter
15	105.5	5.6	843	2	AC2507	hypothetical prote
16	105	5.6	598	2	AB1236	internalin protein
17	104.5	5.5	1441	2	B88807	hypothetical prote
18	103.5	5.5	691	2	F75622	hypothetical prote
19	103.5	5.5	1269	2	A90267	proteinase related
20	103.5	5.5	2399	2	H71879	toxin-like outer m
21	102	5.4	692	2	G90284	hypothetical prote
22	101	5.4	1238	2	AH0038	probable exported
23	100	5.3	1752	2	T48965	hypothetical prote
24	99.5	5.3	5627	2	C83339	hypothetical prote
25	99	5.2	802	2	AH1580	phenylalanyl-tRNA
26	99	5.2	856	2	B81399	probable periplasm
27	99	5.2	1849	2	C41859	Iga-specific metal
28	98.5	5.2	608	2	H64473	hypothetical prote
29	98.5	5.2	4919	2	T31105	hypothetical prote

ALIGNMENTS

RESULT 1

D56617

cfab protein precursor - Escherichia coli plasmid NTP113

C:Species: Escherichia coli

C:Date: 05-Jan-1996 #sequence_revision: 05-Jan-1996 #text_change 10-Dec-1999

C:Accession: D56617

R:Joridi, B.J.; Willshaw, G.A.; van der Zeijst, B.A.; Gastra, W.

DNA Seq. 2, 257-263, 1992

A:Title: The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of hum

A:Reference number: A56617; MUID:9232981; PMID:1352712

A:Accession: D56617

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-360 <J03>

A:Cross-references: GB:M55661; NID:G145507; PIDN:AAC41417.1; PID:G145511

A:Experimental source: enterotoxigenic strain, CFA/I-SI plasmid NTP113

A:Note: sequence extracted from NCBI backbone (NCBI-N:138960, NCBI-P:108972)

C:Genetics:

A:Gene: cfab

A:Genome: plasmid

C:Superfamily: Escherichia colonizing factor antigen cfabE

Query Match	82.6%	Score 1558.5	DB 2	Length 360
Best Local Similarity	81.2%	Pred. No. 1.4e-109		
Matches 293	Conservative 31	Mismatches 36	Indels 1	Gaps 1
Qy	1	MNKLIFITLFFSVLTFPAVSADKIPQESITNLFGPRDRNESSPKNIINNHITAYSE	60	
Db	1	MNKLIFITLFFSSGFFTFVAVSADKIPGSENNTNTIGPHDRGGSSPIYNILNSYLTAYNG	60	
Qy	61	SHTLYDRMTFCLSSHNITNGACPTSENPSSSVSGSETNITLQFTKESLTKRELIQKY	120	
Db	61	SHLYDRMSFTCLSSQNTLNGACPSDDAPGTATIDGETNITLQFTKESLTKRELIQKY	120	
Qy	121	KQLFPKSYNCSGLTNSAHFNCKNKAASGLSYIYPAGELKNLPQGIWDATLKLRYK	180	
Db	121	KQFLFNANCPSKALNGSHFQCNREQASGATLSLYIPAGELKNLPFGGVNVLKLVK	180	
Qy	181	RRYSSTYTYTINITIKLTDKNTQIWLPOKSDARVDNLNRPFGGTYIGNSVDMCFY	240	
Db	181	RRYDTYTYTINITINVTLDKNTQIWLPOKSNARVDNLNRPFGGTYIGNSVDMCFY	240	
Qy	241	DGYSTNSSLEIRFODNPKSDGKPYLRKINDTKEIAYTLLSLAGKSLPTNGTSLNI	300	
Db	241	DGYSTNSSLEIRFODDNDKSDGKPYLRKINDDSKELVYTLSSLLAGKSLPTNGQALNI	300	
Qy	301	ADAASLETNNKRIITAVTMPEISVPVLCWPGRLQDAKVENPEAGQYMGNIINVTTPSSQT	360	
Db	301	-NTASLETNNNRITAVTMPEISVPVLCWPGRLQDAKVENPEAGQYMGNIKITFTPPSSQT	359	
Qy	361	L 361		

Db 360 L 360

RESULT 2

S49539

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S49539

R:Proehlich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A:Title: CoC and CoD are required for assembly of CS1 pili.

A:Reference number: S49539; MID:9434028; PMID:7515003

A:Accession: S49539

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <PRO>

A:Cross-references: EMBL:X76908; NID:G488735; PIDN:CAA54230.1; PID:G488737

C:Superfamily: Escherichia colonizing factor antigen cfaE

Query Match

Best Local Similarity 50.3%; Score 949.5; DB 2; Length 363;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

QY

3 KILFIPTLFFSVLFTFAVSADKIPGDE--STNIP-GER-DRNBSSPKHNLNNHITAY 58

Db

2 KKIFIFL----SIITSVVVSAGRYPETTVGNLTKGFQAPRLDRSQSPYINIFTNHVAGY 57

QY

59 SESHTLYDRMTFLCLSSSHNTLNGACPTSENPSVSSVS--GETNITLQTEKRSLLIKRELOI 117

Db

58 SLSSHLYDRIVFLCTSSSPVNGACPTI---GTSGVQYQYTTTTITLQTEKRSLLIKRNKL 114

QY

118 KGVKLLPKSVNCPG--LTLASAHFNCKNA--ASGASLYLIPAGELXNLPFGGIWDAT 174

Db

115 AGNKPIWNGQCDRSNLVMSKSWSCAHNGAKTLLNLIYPAGEINKLPFGGIWEAT 174

QY

175 IKLRVRRYSET----YGTITNITIKLTDKGNIQIWLPOKSDARVDNLNRPFGGTVI 230

Db

175 LILRLS-RYGEVSSSTHYGNYTNITVDLTDKGNIQIWLPGPHSNPRVDNLNRPIGNKYK 233

QY

231 GNSVDNCFDGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIATLSLLAGKSL 290

Db

234 GNSLDNCFDGYSTNSDSMVIKFQDNDNTNSENLYKIG-GTEKLPYAVSLMGEKIF 292

QY

291 TPNGTSLNIAAASLETWNRTITAVTPEISVPVLCWPGRIQLDAKVENPAGQVWMI 350

Db

293 YPNGOSFTINDSSVLETNWNRVAVAMEVNVVPCWPARILLADVNAPDAGQYSGOI 352

QY

352 NVITFPSSQTL 361

Db

353 YITFPSEVNL 363

RESULT 3

S57937

C:CoD protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S57937

R:Proehlich, B.J.; Karakashian, A.; Melsen, L.R.; Scott, J.R.

submitted to the EMBL Data Library, January 1995

A:Description: The genes for CS2 pili of enterotoxigenic Escherichia coli and their inte

A:Reference number: S57934

A:Accession: S57937

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <PRO>

A:Cross-references: EMBL:Z47800; NID:G897725; PIDN:CAA87763.1; PID:G897729

C:Genetics:

A:Start codon: TGG

C:Superfamily: Escherichia colonizing factor antigen cfaE

Query Match

Best Local Similarity 49.3%; Score 930.5; DB 2; Length 364;

Best Local Similarity 48.4%; Pred. No. 2.1e-62;

Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

QY

1 MKLFIPTLFFSVLFTFAVSADKIPGDESIITNIFGPRDN--ESSPKHNLNNHITAY 58

Db

1 MKKIVFLSKFLCQVYQSWHTVVEAGSINKTSIGPIDRSAAASYFAHVFHEHVAGY 60

QY

59 SESHTLYDRMTFLCLSSSHNTLNGACPTSENPSVSSVSGETNITLQTEKRSLLIKRELOI 118

Db

61 NKDHSLEPRMTFLCWSSTDSKAGACPTGSKSS--QGETN-KLIFTEKSKLARKTLNLK 118

QY

119 GYKQLLFASVNC---PSGLTNSAHFNCKNAASGASLYLIPAGELXNLPFGGIWDATL 175

Db

119 GYKGFLEYSDRCIHYVDKWNLSHTVKCVGSFTGVDFTLIPQGEIDGLTGGWEATL 178

QY

176 KLRVRRYSETVGTVTNITIKLTDKGNIQIWLPOKSDARVDNLNRPFGGTVIGNSV 235

Db

179 ELRVKRRHYDHYGNYKVNITVDLTDKGNIQWTFKFSDDPRIDLNLREGENKYSNVL 238

QY

236 DMCFYDGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIATLSLLAGKSLTPNG 295

Db

239 EMCLYDGYSTHSQSIEMRFQDSSQTNNEYNLIKTEBPLKPLPYKLSLLLGREFYPNG 298

QY

296 TSLNIADAASETWNRTITAVTPEISVPVLCWPGRIQLDAKVENPAGQVWMI 355

Db

299 EAFITNDTSSLFINWRKSVSLPQISIPVLCWPAFLTMSLNPNPEAGEYSGLNVTFT 358

QY

356 PSSQTL 361

Db

359 PSSSL 364

RESULT 4

A30541

probable fimbrial protein tofD [imported] - Salmonella enterica subsp. enterica serovar 7

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: A30541

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: A30502; MUID:21534947; PMID:11677608

A:Accession: A30541

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:G16501589; GSPDB:GN00176

C:Genetics:

A:Gene: tofD

Query Match

Best Local Similarity 14.6%; Score 275.5; DB 2; Length 359;

Matches 90; Conservative 29.7%; Pred. No. 3.2e-13;

Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

QY

73 FLCSSHTLNGACPTSENPSVSSVSGETNITLQTEKRSLLIKRELOIKGYKQLLFKSVN 129

Db

73 WVCNRNENEGACETHLVWVYAGVASKTLRFRQISHAEITL-----ILLGSVR 131

QY

130 --CPGGLTNSAHFNCKNAAS--GASLYLIPAGELXNLPFGGIWDATLKL-RVKRRY 183

Db

132 DACTYGV-----INNMAACQWGRSLKLRIPSEELAKIPTSGTWKATLVLDLWONG 183

QY

184 STYGTGTYTINITIKLTD--KGNIQIWLPOF-KSDARVDNLNRPFGGTVIGNSVDMCY 240

Db

184 DDPFGTSTTDTITLVNTDHFENAAIYFPQFATPRVDNLNHRMASQMSGRANLDMCLY 243

QY

241 DQSYNSSSLEIRFQDNPKSDGKFLRKINDDTKEIATLSLLAGKSLTPTNGTSLNI 300

Db

244 DG-GVKARSLQMKIEGSKNSGTGFQVKSADSAT--IDYAVSMNYGGRSIPVTRGVFSL 300

A:Molecule type: DNA
A:Residues: 1-650 <DEH>
A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60926.1; PID:g854577
R:de Haan, M.; Grievell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56771
A:Accession: S57019
A:Molecule type: DNA
A:Residues: 1-650 <ZAG>
A:Cross-references: EMBL:Z49504; NID:g1015625; PIDN:CAA89526.1; PID:g1015626; MIPS:YJR00
C:Genetics:
A:Gene: SGD:SAG1; AGAL1; AGALPHAI
A:Cross-references: SGD:S0033764; MIPS:YJR004c
A:Map position: 10R
C:Keywords: glycoprotein

Query Match 6.0%; Score 114; DB 2; Length 650;
Best Local Similarity 20.9%; Pred. No. 0.95;
Matches 96; Conservative 52; Mismatches 147; Indels 172; Gaps 23;

QY 4 ILFIPLTFSSVL-----FTFA-----VSADKIPG-----DESTINIFGRDRNE--- 43
DB 8 ILWLFLSALASAININDIFSNLEITPLTANKQPDQGTATDFSTADASSIREGDEFTL 67
QY 44 SSP---KRNILNHIITA-----YSESHLYDRMTFLC-----LSSHNTLN 80
DB 68 SMFHVYRIKLLNSSQAT:SLADGTEAFKQYVSQAAIYENTFTCTAQNDLSSYNTID 127
QY 81 GACPTSENSSSVSGETKITLQFTKRSILKRELOIKGYQLFKSVNC-PSGLTLNSA 139
DB 128 GSITFLNFDGSSVSEYL-----ENAKFKSGPMVLKGMQSDVDFDPAFTENVF 182
QY 140 HFNCNXXAASGASLYL--YIPAGELKNLPFGGIWDATLKLVRKRYSEYTGVTINITIK 197
DB 183 HSGRSTGYSGFSYHLGMYCNGYF---LGG---TEKI-----DYDSSNNVDL- 225
QY 198 LIDKGNIOI-----WLQPKSDARVDLNLPRPGGTIYGRN---SVDMCFYDG--- 242
DB 226 --DCSSVOVYSSNDPNDMFPQSYNDTADV-----TCFGSNLWITLDEKLYDGEML 275
QY 243 -----YSTNSSSLE--IRPDNN--PKDGRKF 265
DB 276 WYNALQSLPANVWIDHAEFFQVTCIDTANTYATQFSTRFRIYQGRNLGTAASKS 335
QY 266 YLRKINDYKXI---AYTL-----LLAGKSLTPNGTSLNTADAA 304
DB 336 FIETTTDLTSLNTSAYSTGSIETVGNRTTSEVISHVVTGKLSPTATTSLT-AQTS 395
QY 305 SLETNW-----RITATMPEISVPVLCWQGR 332
DB 396 IYSTDSNITVGTDLHTTSEVISDVETISRETASTVVAAPTSTTGTGAM 444

RESULT 8
D64962
probable membrane protein b1978 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: D64962
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: D64962
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2383 <BLAT>
A:Cross-references: GB:AE000289; GB:U00096; NID:gl788285; PIDN:AACT5042.1; PID:gl788289;
A:Experimental source: strain K-12, substrain MG1655
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:54-70/Domain: transmembrane #status predicted <TM>
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 5.8%; Score 109; DB 2; Length 2383;
Best Local Similarity 20.0%; Pred. No. 12;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLTFEAVSADKIPQDESITNIFGPRORNESSPXNINLNEITAY---SESHLYDRMT 69
DB 762 SAKIATLSASNGVLANENANTVSVNVADEGS---NPINDETVTFAVLSGATSENNON 818
QY 70 FCLSSHNTLNGACPTSENPSVSSVGETNITLQFTKRSLI-----K 112
DB 819 ---TAKTDVNGLA-TFDLKSSKQEDNTVEVITLNGVKOTLIVSFVGDSSTAQVDLQSK 873
QY 113 RELQIKGYKQL-LFKSVNCPGSLTNSHPNCKNAASGASILYIPAGELKNLPFGGIW 171
DB 874 NEVVADGNDSVTMTATVDAKGNLLNDVMVTFNVASAEAKLSQTEVNSHD-----GIA 926
QY 172 DATLKLVRKRYSEYTGVTINITIKLTDKGNIQWLQPKSDARVDLMLRPTGGGTIY 231
DB 927 TAILT-----SLKNGDYRVTSVSSSQANQVNFIDQSTAAITLSV-PSGDTIV-- 976
QY 232 RNSVDVCFYDGYSTRSS-----SLEIRFQDNNPKSGKFYLRKINDDKETIATLSLLAG 287
DB 977 -----TNTAPQYMTATLQDKNGNPLKDKBITFSVPND---VASKFSISNG 1019
QY 288 KSLTPTNGTSL-----NIADAASLETNWNRLITAYTMPEISVPVLC 327
DB 1020 KGMTDSNGVALASLTGTLAGTHMIRANLSVSDAQPMTFVADKRVVVLQTSKAEII 1079
QY 328 WFG--RLQLDARVENPEAGQYMGINVTFT 355
DB 1080 GNGVDETILTATVKOP-SNHPVAGITVNET 1108

RESULT 9
S55097
probable membrane protein YMR215w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR261.09
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S55097
R:Dedman, K.; Brown, D.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55089
A:Accession: S55097
A:Molecule type: DNA
A:Residues: 1-524 <DED>
A:Cross-references: EMBL:Z49809; NID:g854459; PID:g854467; GSPDE:GN00013; MIPS:YMR215w
A:Experimental source: strain AB972
C:Genetics:
A:Gene: MIPS:YMR215w
A:Cross-references: SGD:S0004828
A:Map position: 13R
C:Superfamily: glycopospholipid-anchored surface glycoprotein GAS1
C:Keywords: transmembrane protein
F:7-23/Domain: transmembrane #status predicted <TM>

Query Match 5.7%; Score 108; DB 2; Length 524;
Best Local Similarity 24.7%; Pred. No. 2;
Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

QY 119 GYKQL--LFKSVNCPGSLTNSAHFNCKNAA---SGASILYIPAGELKNLPFGGIWDA 173
DB 264 GYDKLNSTFEDAVIP---LIFSEYCNKNTPTTFDEVESEGLY---GGLNVPFSGSL--- 313
QY 174 TLKLVRKRYSEYTGVTINITIKLTDKGNIQWLQPKSD-ARVDLNLR-----PTGGGT 228
DB 314 -----VVEYTEANNYGL---VKLDDSGSL-----TYKDDFVNLSQLKNVSLPTKES 359
QY 229 YIGRNSVDWC-----FYDGYSTNSSSLE-----IRFQNNPKSGDK----- 264
DB 360 EISSDSIYKCDNSAITN-YSGFTNNTFLPSQPAETANNMIEYGVNGTNT-GKILTDYAVP 418

265 ----PYLRKINDDTKEIATYLSLLLAGKSLTPTNGFSLNIAADAASLETNWAR-ITAVTMP 319
DB 419 TTFNYTIKNNKDDT--ISATISYDKA-NSLNELDVTATTVAASASTSSSSSLTSSSTSP 475
320 ET-SVPVLCWFGRLQLDKVENPEAGYMGNI-NVTFTPS 357
DB 476 SSTGSSSSSTG-----SSSASSSSKSGVGNIVNVFSQS 510

RESULT 10
B64635
toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: B64635
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 386, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Praser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: B64635
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2529 <TOM>
A:Cross-references: GB:AE000602; GB:AE000511; NID:92314060; PIDN:AD07969.1; PID:G231406

Query Match 5.7%; Score 108; DB 2; Length 2529;
Best Local Similarity 22.5%; Pred. No. 16;
Matches 84; Conservative 50; Mismatches 122; Indels 118; Gaps 22;

QY 85 TSENTESSSVSGETNITLQFTKRSLL-----KRELQIKYKQLFK--SVNCP 131
DB 977 TSNFNATTQLGNTNFTL--SSQSLNFNGDPTLQNNANITLGNKSQAFFKNSLTLDNN 1033

QY 132 SGTLL-NSAFENCKNAA--SGASLYLYIPA-GEKLNLPFGGIWDATLRLVKRRYSY 187
DB 1034 SNLSLDNQSVLANNTSAPFNKQASLNYNGSQATFNSLPFNG---GTLSLNASSKLNAGN 1090

QY 188 GYVITNITIKLTD-----KGNIGIWLFPQFKSDARVDLNLPTGGGYVIGRNSVD 236
DB 1091 AFSFNNTNINLDDSVLSASNTSLNANI-----NFGASQADF-----GNTIITDASEN 1140

QY 237 MCFYGYSTNSSL-----BIRFOONPK-----SDGKFYLRKIN-----DDT 274
DB 1141 --FDSASSLNFNNTANGALNFNGYTPSLTKALMSVSGQFVLNNGDINLSDINIFDNIT 1198

QY 275 KEIAYTSLILLAGKSLTPTNGTS-----LNIAADASLETN---WNRTITAVTMPEI 321
DB 1199 KSVTY--NIINAKGKITGISGANGYKILFYGMKIQNATYSDNNNIQTWSFNPINSSQI 1256

QY 322 -----SVEVL-----CWFGRLQLDAKVENP-----EAGQY--M 347
DB 1257 IQBSIKNGDLTIEVNPNSASNTIFENIAPELNYQASKNQKPGYSYSDNQAGTYILT 1316

QY 348 GNINVTFTTP-SSOT 360
DB 1317 SNIKGLFTPKGSOT 1330

RESULT 11
A64904
probable fimbrial protein b1502 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64904
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, K.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64904

A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-304 <BLAT>
A:Cross-references: GB:AE000247; GB:U00096; NID:G1787773; PIDN:AAC74575.1; PID:G1787779;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: fimbrial protein film
C:Keywords: fimbria

Query Match 5.7%; Score 107; DB 2; Length 304;
Best Local Similarity 21.0%; Pred. No. 1.2;
Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IFTLFPSSVLFTFAVSADKIPGDESI---TNIFGPRDRNESSPKHNI---LNNH 54
DB 3 KVLFGIYLLLMAGKVAFSCNVD---GGSSICAGTTSVYVNLD-PVIQGNLVVDLSOH 63

QY 55 ITAYSESHLYDRMTFLCLSSH-NTLNGACPTSENPSSSVSGETNITLQFTKRSLLKR 113
DB 64 ISCWNDYGGWYD-----TDHINLVQG-----SAFAG----- 89

QY 114 ELQIKGYK-QLLFKSVNCPSEGLTNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIWD 172
DB 90 --SLQSYKGSLYMNVVTFPPLTTNTNVLIDIGKTPMLPLKLYI-----TPVGAAG 140

QY 173 ATLK-----LRVKRRYSYETG-----YTINITIK-----LTDKGNIOIWLPOPK 212
DB 141 VIKAGEVIARIHMTKIALGSGNPNFTWNIISNNVNVPTGGCTVDSRNTVLDLPDP 200

QY 213 SPARVDLNLRLPTGGGYTIGRNSVDMCFYGYSTNSSSLBIRFQDNNPKSDGKFYLRKIND 272
DB 201 GSAETPL-----GVYCS-SQKLSFY--SGATTDSSROV-FANTAP-----D 239

QY 273 DKETAATLSLLLAGK-----SLTPNGISLNIADAASLETNWNRITAVTMPEI 321
DB 240 ATKASGVGTLMRNGKILATGENVSLGTYNKSKVPLGLSATYGTGCKNKSAGEVQSV 296

RESULT 12
C90892
probable adhesin [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90892
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinaigawa, H. DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90892
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <RAY>
A:Cross-references: GB:BA000007; PIDN:BA035530.1; PID:G13361573; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs2107
C:Superfamily: fimbrial protein filmH

Query Match 5.7%; Score 107; DB 2; Length 304;
Best Local Similarity 21.0%; Pred. No. 1.2;
Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19;

QY 3 KILF-IFTLFPSSVLFTFAVSADKIPGDESI---TNIFGPRDRNESSPKHNI---LNNH 54
DB 8 KVLFGIYLLLMAGKVAFSCNVD---GGSSICAGTTSVYVNLD-PVIQGNLVVDLSOH 63

QY 55 ITAYSESHLYDRMTFLCLSSH-NTLNGACPTSENPSSSVSGETNITLQFTKRSLLKR 113
DB 64 ISCWNDYGGWYD-----TDHINLVQG-----SAFAG----- 89

QY 114 ELQIKGYK-QLLFKSVNCPSEGLTNSAHFNCKNKAASGASLYLYIPAGELKNLPFGGIWD 172
DB 90 --SLQSYKGSLYMNVVTFPPLTTNTNVLIDIGKTPMLPLKLYI-----TPVGAAG 140

Cy	210	QFKSDARDVNL-----LRPT-----GGTGTIGNSVDMCYDGVSTNSSSJE-----	251
		:	
		:	
Db	472	-FIKSNIDISGQSALLRPSNITSTFCHGNGRNDINTLNLIIISNGGGISSSTLSAGKA	530
		:	
		:	
Qy	252	-----IRFQDNPK--SDGKFLRKINDDFKELAYTLLSLLAGKS-----	289
		:	
		:	
Db	531	GNISNGSNINVVGTNINSFINSNPLL--VDPNLKLLYRQDPLLIGQGNIFL	588
		:	
		:	
Qy	290	----LTTNGTSLN-----IADAASLETNNRRIYAVTMPE-SVPVLCWP-GRLOLDARKE	339
		:	
		:	
Db	589	NTDIINISNGGLINARNEGVDAGNIRISANTINISQGEVNATTTIGEGGNIILNSR--	646
		:	
		:	
Qy	340	NPEAGQYMGNINVTFT	355
		:	
		:	
Db	647	----NLEFNNSRITAT	658

Search completed: May 6, 2004, 10:19:44
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 10:15:01 ; Search time 46 Seconds
(without alignments)
2476.132 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFVTFSSVLTFA.....BAGQYMGXNVTFTSSQTL 361

Scoring table: ELOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREXBL.25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriopl.*
17: sp_archaea.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result- being printed, and is derived by analysis of the total score distribution.

SUMMARY35

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2 Q93G67	Q93G67 escherichia
2	950.5	50.4	363	2 Q848J5	Q848J5 escherichia
3	949.5	50.3	363	2 Q47115	Q47115 escherichia
4	930.5	49.3	364	2 Q47119	Q47119 escherichia
5	381	20.2	387	2 Q8KQJ7	Q8KQJ7 burkholderi
6	371	19.7	387	2 Q8KQJ3	Q8KQJ3 burkholderi
7	275.5	14.6	359	16 Q9XDS1	Q9XDS1 salmonella
8	272.5	14.4	245	16 Q83J01	Q83J01 shigella fl
9	256.5	13.6	210	16 Q7UBH8	Q7UBH8 shigella fl
10	127	6.7	1881	16 Q8RGK2	Q8RGK2 fusobacteri
11	115.5	6.1	700	2 Q8VU50	Q8VU50 chlamydia p
12	115.5	6.1	847	2 P71L32	P71L32 chlamydophi
13	114	6.0	865	15 Q8Q7H7	Q8Q7H7 human immun
14	113	6.0	2646	5 Q8I521	Q8I521 plasmodium
15	111.5	5.9	743	16 Q8XNW2	Q8XNW2 clostridium
16	111	5.9	2768	16 Q8ESG6	Q8ESG6 shewanella

17	110.5	5.9	589	16 Q3F994	Q8F994 leptospira
18	110.5	5.9	1052	17 Q8QNS0	Q8QNS0 methanocarc
19	110	5.8	265	2 Q8KGI5	Q8KGI5 rhizobium l
20	110	5.8	846	2 P71L33	P71L33 chlamydophi
21	109.5	5.8	2454	5 Q86B02	Q86B02 dictyosteli
22	109	5.8	807	5 Q8T7V5	Q8T7V5 drosophila
23	109	5.8	843	16 Q823X3	Q823X3 chlamydophi
24	108	5.7	872	2 Q9RLA0	Q9RLA0 rickettsia
25	108	5.7	942	16 Q823X1	Q823X1 chlamydophi
26	108	5.7	2529	16 Q25579	Q25579 helicobacte
27	107	5.7	304	16 Q9XAX2	Q9XAX2 escherichia
28	106.5	5.6	1012	16 Q8PEE7	Q8PEE7 xanthomonas
29	106.5	5.6	1065	16 Q8EZX0	Q8EZX0 leptospira
30	106	5.6	1937	5 Q8IE94	Q8IE94 plasmodium
31	105.5	5.6	430	3 Q96VN2	Q96VN2 blumeria gr
32	105.5	5.6	943	16 Q8YKQ8	Q8YKQ8 arabidopsis
33	105	5.6	598	16 Q8Y7I7	Q8Y7I7 listeria mo
34	105	5.6	868	16 Q823X4	Q823X4 chlamydophi
35	105	5.6	2931	5 Q86AE3	Q86AE3 dictyosteli
36	104.5	5.5	1441	16 Q9CPL1	Q9CPL1 lactococcus
37	104	5.5	304	16 Q5CW27	Q5CW27 escherichia
38	104	5.5	803	10 Q8AUS5	Q8AUS5 oryza sativ
39	103.5	5.5	691	16 Q9ZS7	Q9ZS7 deinococcus
40	103.5	5.5	1269	17 Q97Z06	Q97Z06 sulfolobus
41	103.5	5.5	2399	16 Q9ZKX9	Q9ZKX9 helicobacte
42	103	5.5	868	17 Q8PZM5	Q8PZM5 methanocarc
43	103	5.5	1754	16 Q8D81	Q8D81 thermococcus
44	102.5	5.4	1736	5 Q85PH7	Q85PH7 dictyosteli
45	102.5	5.4	1736	5 Q86A19	Q86A19 dictyosteli

ALIGNMENTS

RESULT 1

Q93G67 PRELIMINARY; PRT; 361 AA.
AC Q93G67; 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Minor pilin protein CsaE.
GN CsaE.
CS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E11881A;
RA Altboum Z.D., Levine M.M., Galen J.E., Barry E.M.;
RT "Isolation and Characterization of ETEC C84-imbriae encoding genes,
RL and their expression in Shigella flexneri 2a guaba strain CVD 1204.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF296132; AAK97137.1; -
SQ SEQUENCE 361 AA; 4C102 MW; 2E1E74ABD80EB353 CRC64;

Query Match	100.0%;	Score	1886;	DB	2;	Length	361;
Best Local Similarity	100.0%;	Pred. No.	6.4e-135;				
Matches	361;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MNKLIFLTFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE	60				
Db	1	MNKLIFLTFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE	60				
QY	61	SHLYDRMTPLCSSLNLTNGACPTSENSSSVSGETNITLQFTTKRSLSIKELQIKGY	120				
Db	61	SHLYDRMTPLCSSLNLTNGACPTSENSSSVSGETNITLQFTTKRSLSIKELQIKGY	120				
QY	121	KQLLFKSVNCPSGITLNSAHFNCKNAASCASLYLYIPAGELKNLPFGGWDATLKVVK	180				
Db	121	KQLLFKSVNCPSGITLNSAHFNCKNAASCASLYLYIPAGELKNLPFGGWDATLKVVK	180				

```
QY 181 RRYSETGYTNTITIKLTDKGNIQIWLPOKSDARVDNLNRPRTGGTGYGRNSVDMCFY 240
DB 181 RRYSETGYTNTITIKLTDKGNIQIWLPOKSDARVDNLNRPRTGGTGYGRNSVDMCFY 240
QY 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSLTPNGTSLNI 300
DB 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSLTPNGTSLNI 300
QY 301 ADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNIINVTFPSSQT 360
DB 301 ADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNIINVTFPSSQT 360
QY 361 L 361
DB 361 L 361
RESULT 2
ID Q848J5 PRELIMINARY; PRT; 363 AA.
AC Q848J5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CSDD.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC plasmid unnamed.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS37-4;
RA Biediger W., Wolf M.K.;
RT "CS17 of Enterotoxigenic Escherichia coli.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY216431; AAC60098.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 363 AA; 40145 MW; 0911574F0406C54A CRC64;
Query Match 50.4%; Score 950.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 5.4e-64;
Matches 197; Conservative 54; Mismatches 99; Indels 21; Gaps 11;
QY 3 KILFIPLTFSSVLTFAVSGADKIPGDE--SITNIF-GPR-DRNESSPKHILNHHITAY 58
DB 2 KKIPIFL----SIIFSVAWSAGRYPETTVGNLTGKSFQAPLDRSVQSPYINFTNHVAGY 57
QY 59 SESHTLYDRMTFLCLSSHTNLGACPTSENPSSSVS-GETNITLQFTEKRSLIKRELQI 117
DB 58 SLSHSLYDRIVFLCTSSSNPNVGACP--NRGTSGVQYGTITITLQFTEKRSLIKRNINL 114
QY 118 KGYKQLLFKSVNCPG--LTNSAHFNCNKA-ASGASLYIYPAGELKNLPFGGIWDAT 174
DB 115 AGNKKPIWENQSCDTSNLMVLNKSWSGANGNANGTLNLYIPAGEINKLPFGGIWEAT 174
QY 175 LKLRVKRYSYSET---YGYTITITIKLTDKGN:QIWLPOKSDARVDNLNRPRTGGGYI 230
DB 175 LILRLS-RYGEVSTHYGNYTAVITVDLTDKGN:QVWLPGFHSNPRVDNLNRPRTGNKYK 233
QY 231 GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSL 290
DB 234 GSNLDMCFYDGYSTNSDXYIKQDDNPTYSSEVNLKIG-GTEKLPYAVSLLMGEKIF 292
QY 291 TPTNGTSLNIADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNI 350
DB 293 HPVNGQSTINDSSVLETNWNRVTAVAMPVNVPLVCPARLLNADVNADPAGQYSGQI 352
QY 351 NVTFPSSQTL 361
DB 353 YITFTPSVENL 363
RESULT 3
ID Q47115 PRELIMINARY; PRT; 363 AA.
AC Q47115;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cood precursor.
GN Cood.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94344028; PubMed=7915003;
RA Froehlich B., Karakashian A., Melsen L.R., Wakefield J.C., Scott J.R.;
RT "Cooc and Cood are required for assembly of CSI pilli.";
RL Mol. Microbiol. 12:387-401(1994).
DR EMBL; X76908; CAA54230.1; -.
DR PIR; S49539; S49539.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0004812; P:RNA ligase activity; IEA.
DR GO; GO:0006418; P:amino acid activation; IEA.
DR InterPro; IPR001412; rRNA-synt 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Signal.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 18 POTENTIAL.
SQ SEQUENCE 363 AA; 40139 MW; AA9488AADDSD5F72 CRC64;
Query Match 50.3%; Score 949.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 6.4e-64;
Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;
QY 3 KILFIPLTFSSVLTFAVSGADKIPGDE--SITNIF-GPR-DRNESSPKHILNHHITAY 58
DB 2 KKIPIFL----SIIFSVAWSAGRYPETTVGNLTGKSFQAPLDRSVQSPYINFTNHVAGY 57
QY 59 SESHTLYDRMTFLCLSSHTNLGACPTSENPSSSVS-GETNITLQFTEKRSLIKRELQI 117
DB 58 SLSHSLYDRIVFLCTSSSNPNVGACP--NRGTSGVQYGTITITLQFTEKRSLIKRNINL 114
QY 118 KGYKQLLFKSVNCPG--LTNSAHFNCNKA-ASGASLYIYPAGELKNLPFGGIWDAT 174
DB 115 AGNKKPIWENQSCDTSNLMVLNKSWSGANGNANGTLNLYIPAGEINKLPFGGIWEAT 174
QY 175 LKLRVKRYSYSET---YGYTITITIKLTDKGN:QIWLPOKSDARVDNLNRPRTGGGYI 230
DB 175 LILRLS-RYGEVSTHYGNYTAVITVDLTDKGN:QVWLPGFHSNPRVDNLNRPRTGNKYK 233
QY 231 GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSL 290
DB 234 GSNLDMCFYDGYSTNSDXYIKQDDNPTYSSEVNLKIG-GTEKLPYAVSLLMGEKIF 292
QY 291 TPTNGTSLNIADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNI 350
DB 293 YPVNGQSTINDSSVLETNWNRVTAVAMPVNVPLVCPARLLNADVNADPAGQYSGQI 352
QY 351 NVTFPSSQTL 361
DB 353 YITFTPSVENL 363
RESULT 4
ID Q47119 PRELIMINARY; PRT; 364 AA.
AC Q47119;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
QY 181 RRYSETGYTNTITIKLTDKGNIQIWLPOKSDARVDNLNRPRTGGTGYGRNSVDMCFY 240
DB 181 RRYSETGYTNTITIKLTDKGNIQIWLPOKSDARVDNLNRPRTGGTGYGRNSVDMCFY 240
QY 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSLTPNGTSLNI 300
DB 241 DGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSLTPNGTSLNI 300
QY 301 ADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNIINVTFPSSQT 360
DB 301 ADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNIINVTFPSSQT 360
QY 361 L 361
DB 361 L 361
RESULT 2
ID Q848J5 PRELIMINARY; PRT; 363 AA.
AC Q848J5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CSDD.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC plasmid unnamed.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS37-4;
RA Biediger W., Wolf M.K.;
RT "CS17 of Enterotoxigenic Escherichia coli.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY216431; AAC60098.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 363 AA; 40145 MW; 0911574F0406C54A CRC64;
Query Match 50.4%; Score 950.5; DB 2; Length 363;
Best Local Similarity 53.1%; Pred. No. 5.4e-64;
Matches 197; Conservative 54; Mismatches 99; Indels 21; Gaps 11;
QY 3 KILFIPLTFSSVLTFAVSGADKIPGDE--SITNIF-GPR-DRNESSPKHILNHHITAY 58
DB 2 KKIPIFL----SIIFSVAWSAGRYPETTVGNLTGKSFQAPLDRSVQSPYINFTNHVAGY 57
QY 59 SESHTLYDRMTFLCLSSHTNLGACPTSENPSSSVS-GETNITLQFTEKRSLIKRELQI 117
DB 58 SLSHSLYDRIVFLCTSSSNPNVGACP--NRGTSGVQYGTITITLQFTEKRSLIKRNINL 114
QY 118 KGYKQLLFKSVNCPG--LTNSAHFNCNKA-ASGASLYIYPAGELKNLPFGGIWDAT 174
DB 115 AGNKKPIWENQSCDTSNLMVLNKSWSGANGNANGTLNLYIPAGEINKLPFGGIWEAT 174
QY 175 LKLRVKRYSYSET---YGYTITITIKLTDKGN:QIWLPOKSDARVDNLNRPRTGGGYI 230
DB 175 LILRLS-RYGEVSTHYGNYTAVITVDLTDKGN:QVWLPGFHSNPRVDNLNRPRTGNKYK 233
QY 231 GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSDGKFLRKINDTKETIAITLSLLAGKSL 290
DB 234 GSNLDMCFYDGYSTNSDXYIKQDDNPTYSSEVNLKIG-GTEKLPYAVSLLMGEKIF 292
QY 291 TPTNGTSLNIADAASLETNNRITAVTMPETISVPLVCPGRQLQDAKVENPEAGQYMGNI 350
DB 293 HPVNGQSTINDSSVLETNWNRVTAVAMPVNVPLVCPARLLNADVNADPAGQYSGQI 352
QY 351 NVTFPSSQTL 361
DB 353 YITFTPSVENL 363
```

DE COTD precursor.
GN COTD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OT Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C91F-6;
RA MEDLINS=96071908; PubMed=7591145;
RA Frechlich B.J., Karakashian A., Sakellaris H., Scott J.R.;
RT "Genes for CS2 pilli of enterotoxigenic Escherichia coli and their
RT interchangeability with those for CS1 pilli";
RL Infect. Immun. 63:4849-4856(1995).
DR EMBL; Z47800; CAA87763.1; -
DR PIR; S57937; S57937.
KW Signal.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 364 COTD.
SQ SEQUENCE 364 AA; 40979 MW; ECTAC732A0E195B6 CRC64;

Query Match 49.3%; Score 930.5; DB 2; Length 364;
Best Local Similarity 48.4%; Pred. No. 1.8e-62;
Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

Qy 1 MNKILFTLFFSVLFTPAVSADKIPGDESITNIFGPRDRN--ESSPKHMLNKHITAY 58
Db 1 MKKVFVLSMFLCSQVYQSWHTNVEAGS--NKTSES--GPIDRSAAASYPARHIFHERVAGY 60

Qy 59 SESHTLYDRMTFLGSSHTNLNGACPTSENFSSSVSGETNITLQFTKRLIKRELIQIK 118
Db 61 NKDHSLEDRMTFLGSSHTNLNGACPTSENFSSSVSGETNITLQFTKRLIKRELIQIK 118

Qy 119 GYKOLLEFVNC--PSGLTILNSAHFNCKNAAGSALYLYIPAGELKVLPPFGIWDATL 175
Db 119 GYKRLFYEDRCIHHVDKMLNSHCKVCGSFTGVDFTLYPQGEIDGLLTGGIWEATL 178

Qy 176 KLRVKRRSYETGYTINITIKLTDKGNIOIWLPOFKSDARVDNLNRPETGGTY--GRNSV 235
Db 179 ELRVKRRHYDNGHYTKVNTVDLTDKGNIQW--FKFHSRPRIDNLNRPETGGTY--GRNSV 238

Qy 236 DMCFYDGYSTNSSLFIRFQDNPKSDGKLYRKINDTKKIAIYTLISLLAGKSLTPNG 295
Db 239 EMCLYDGYSTNSSLFIRFQDNPKSDGKLYRKINDTKKIAIYTLISLLAGKSLTPNG 298

Qy 296 TSLTADAASLETNWRITAVTMPEISVPLVCPQRLQDLAKVENPEAGQYMGNIIVFT 355
Db 299 EAPITNDTSSSLFNNRIRKISVSLPISIPVLCWPAULCFMSEINNPEAGYSGILNVFT 358

Qy 356 PSSQTL 361
Db 359 PSSSSL 364

RESULT 5
Q8KQJ7 PRELIMINARY; PRT; 387 AA.
ID Q8KQJ7
AC Q8KQJ7
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative minor pilin protein.
GN CBLD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BC7;
RA Tomich M., Mohr C.D.;
RT "Identification and Characterization of a locus required for cable
RT pilus biogenesis in Burkholderia cepacia";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV114293; AA56040.1; -
DR InterPro; IPR003016; LipovL BS.
DR PROSITE; PS00189; LIPOYL; 1-
SQ SEQUENCE 387 AA; 41805 MW; 39DC45F31EA8CE0C CRC64;

Query Match 19.7%; Score 371; DB 2; Length 387;
Best Local Similarity 31.7%; Pred. No. 5e-20;
Matches 114; Conservative 50; Mismatches 144; Indels 52; Gaps 13;

Qy 39 RDRNESPKNILNKHITAYSSHTL-YDRMTFLGSSHTNLNGACPTSENPSSSSVSSE 97
Db 43 RDRGTPVGDVLIIDFKASGGYDTSYLVKWRNSWTQCSSTDTATGACPTV--FVWEKAGTS 100
Qy 98 TNLTLQFTKRLIKRELIQIK--GYKO-----LLFKSVNCP--SGLTILNSAHFN 144
Db 101 TNLKLSFKENTGATAVLTLEGANGYORHRDCISVSGVGVDCPKTDENESATAGCD 160
Qy 145 KNAASGASLYIYPAGELKVLPPFGIWDATLKRVKRRSYETGYTINITIKLTDK 201
Db 161 TVGWDGRVLFVKIPAAELKKLPSSGTWKANLRLNKLWSSSTATTALFKAATILDTDK 220
Qy 202 GNIOIWLPOFKS--DARVDNLNRPETGGTY--GRNSVDMCFYDGYSTNSSLFIRFQDN 257
Db 221 NNIOQVILPEFTSATPTVDLKLNRVANGSRMGSVSNVDVCLYDGYNSQSTWPDVSASDGLT 280
Qy 258 -NPKSDGKLYRKINDTK--XKIAIYTLISLLAGKSLTPNGTSLNIAADAASLETNWR 312
Db 281 IDRRDKQYSTLLDKRSGAYESIDVAASLYAGKIALFNENETVRLQG-----VNSA 335
Qy 313 ITAVTMPEISVPLVCPQRLQDLAKVENPE-----AGQYMGNIIVFTPSSQTL 361
Db 336 GRSVTLPGISAPVICTPTFLTL----ETPEFQSVKMPGKYSHKLTITFTPSSTSL 387

RESULT 6
Q8KQ10 PRELIMINARY; PRT; 387 AA.
ID Q8KQ10
AC Q8KQ10;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative minor pilin and initiator.
GN CBLD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BC7;
RA Tomich M., Mohr C.D.;
RT "Identification and Characterization of a locus required for cable
RT pilus biogenesis in Burkholderia cepacia";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV114293; AA56040.1; -
DR InterPro; IPR003016; LipovL BS.
DR PROSITE; PS00189; LIPOYL; 1-
SQ SEQUENCE 387 AA; 41805 MW; 39DC45F31EA8CE0C CRC64;

Query Match 19.7%; Score 371; DB 2; Length 387;
Best Local Similarity 31.7%; Pred. No. 5e-20;
Matches 114; Conservative 50; Mismatches 144; Indels 52; Gaps 13;

Qy 39 RDRNESPKNILNKHITAYSSHTL-YDRMTFLGSSHTNLNGACPTSENPSSSSVSSE 97
Db 43 RDRGTPVGDVLIIDFKASGGYDTSYLVKWRNSWTQCSSTDTATGACPTV--FVWEKAGTS 100
Qy 98 TNLTLQFTKRLIKRELIQIK--GYKO-----LLFKSVNCP--SGLTILNSAHFN 140
Db 101 TNLKLSFKENTGATAVLTLEGANGYORHRDCISVSGVGVDCPKTDENESATAGCD 156
Qy 141 FNCNKNAAGSALYIYPAGELKVLPPFGIWDATLKRVKRRSYETGYTINITIK 197
Db Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY982893; AAL92876.1; -
DR InterPro; IPR003016; LipovL BS.
DR PROSITE; PS00189; LIPOYL; 1-
SQ SEQUENCE 387 AA; 41653 MW; B4547E146DDAC7B2 CRC64;

Query Match 20.2%; Score 381; DB 2; Length 387;
Best Local Similarity 32.6%; Pred. No. 8.7e-21;
Matches 116; Conservative 50; Mismatches 146; Indels 44; Gaps 13;

Qy 39 RDRNESPKNILNKHITAYSSHTL-YDRMTFLGSSHTNLNGACPTSENPSSSSVSSE 97
Db 43 RDRGTPVGDVLIIDFKASGGYDTSYLVKWRNSWTQCSSTDTATGACPTV--FVWEKAGTS 100
Qy 98 TNLTLQFTKRLIKRELIQIK--GYKO-----LLFKSVNCP--SGLTILNSAHFN 144
Db 101 TNLKLSFKENTGATAVLTLEGANGYORHRDCISVSGVGVDCPKTDENESATAGCD 160
Qy 145 KNAASGASLYIYPAGELKVLPPFGIWDATLKRVKRRSYETGYTINITIKLTDK 201
Db 161 TVGWDGRVLFVKIPAAELKKLPSSGTWKANLRLNKLWSSSTATTALFKAATILDTDK 220
Qy 202 GNIOIWLPOFKS--DARVDNLNRPETGGTY--GRNSVDMCFYDGYSTNSSLFIRFQDN 257
Db 221 NNIOQVILPEFTSATPTVDLKLNRVANGSRMGSVSNVDVCLYDGYNSQSTWPDVSASDGLT 280
Qy 258 -NPKSDGKLYRKINDTK--XKIAIYTLISLLAGKSLTPNGTSLNIAADAASLETNWR 312
Db 281 IDRRDKQYSTLLDKRSGAYESIDVAASLYAGKIALFNENETVRLQG-----VNSA 335
Qy 313 ITAVTMPEISVPLVCPQRLQDLAKVENPE-----AGQYMGNIIVFTPSSQTL 361
Db 336 GRSVTLPGISAPVICTPTFLTL----ETPEFQSVKMPGKYSHKLTITFTPSSTSL 387

RESULT 6
Q8KQ10 PRELIMINARY; PRT; 387 AA.
ID Q8KQ10
AC Q8KQ10;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative minor pilin and initiator.
GN CBLD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BC7;
RA Tomich M., Mohr C.D.;
RT "Identification and Characterization of a locus required for cable
RT pilus biogenesis in Burkholderia cepacia";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV114293; AA56040.1; -
DR InterPro; IPR003016; LipovL BS.
DR PROSITE; PS00189; LIPOYL; 1-
SQ SEQUENCE 387 AA; 41605 MW; 39DC45F31EA8CE0C CRC64;

Query Match 19.7%; Score 371; DB 2; Length 387;
Best Local Similarity 31.7%; Pred. No. 5e-20;
Matches 114; Conservative 50; Mismatches 144; Indels 52; Gaps 13;

Qy 39 RDRNESPKNILNKHITAYSSHTL-YDRMTFLGSSHTNLNGACPTSENPSSSSVSSE 97
Db 43 RDRGTPVGDVLIIDFKASGGYDTSYLVKWRNSWTQCSSTDTATGACPTV--FVWEKAGTS 100
Qy 98 TNLTLQFTKRLIKRELIQIK--GYKOLLEFVNC-----PSGLTILNSAH 140
Db 101 TNLKLSFKENTGATAVLTLEGANGYORHRDCISVSGVGVDCPKTDENESATAGCD 156
Qy 141 FNCNKNAAGSALYIYPAGELKVLPPFGIWDATLKRVKRRSYETGYTINITIK 197
Db Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

```

Db 157 AGCDTVGVNDGRVLPVKIPAAELKGLPSCGTXKANRLNKLWSSSTATTTLA:FKAAITLD 216
QY 198 LTKGNQIQLWLPQKX-DARVDNLNRPTGGTYI-GRNSVDMCFYDGYSTNSSSLEIRFQ 255
Db 217 TTKXNIQVYLSEFFSATPTVDLKLKPLNGSMGTSNDMLDYGNSQSTWFDVSAS 276
QY 256 DN---NPKSDGDFYLRKINDT---KEIATYLSLLAGKSLTPPTNGTSLNIADAASLET 308
Db 277 DGIITDRKRGQVSTLDDKSGAVESRDYAAASLTFAKXIALPNNETVRLQG-----V 331
QY 309 NNRRITAVTMPEISVPLVCPGRQLQDAKVEP-----AGQYMGNLNVTFPSQTL 361
Db 332 NNSAGRSVTLPGTSAPVICTPFLTL-----ETPEFQSVKRPCKYSHKLTITFTPSSTSL 387

RESULT 7
Q9XDS1 PRELIMINARY; PRT; 359 AA.
AC Q9XDS1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE TSAD protein (putative fimbrial protein).
GN TSAD OR TCFD OR STY0348 OR W2547.
OS Salmonella enterica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601, 28901;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=GIFU10007;
RC Hashimoto Y.;
RA "Identification of a putative fimbrial operon of Salmonella typhi.";
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=RKS 3333;
RC MEDLINE=99348391; PubMed=10417651;
EX Folkesson A., Advani A., Sukupolvi S., Pfeifer J.D., Normark S.,
RA Lofdahl S.;
RA "Multiple insertions of fimbrial operons correlate with the evolution
RT of Salmonella serovars responsible for human disease.";
RL Mol. Microbiol. 33:612-622(1999).
RN [3]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtzoyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail W., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=92531367; PubMed=12644504;
EX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
RN [5]
EMBL; AB029403; BAB82272.1; -
DR EMBL; AJ242964; CAB51577.1; -
DR EMBL; AL627266; CAD08773.1; -
DR EMBL; AE016842; AA070131.1; -
DR PI3; AE0541; AE0541.

```

```

KW Hypothetical protein.; Complete proteome.
SQ SEQUENCE 359 AA; 39741 MW; BB88A4E9A62052CE CRC64;
Query Match 14.6%; Score 275.5; DB 16; Length 359;
Best Local Similarity 29.7%; Pred. No. 7.8e-13;
Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

QY 70 FICLSHNTLNGACSTSNPSSSVSGSTNTYLTQTKESLIKRELOIKGYQLPKSVN 129
Db 73 WVCRRNRNENEGACETHLVWVYAFGAYSKRLRFREQISHAEITL-----ILLGVR 131
QY 133 --CPSSLTNSAHFNCKNAAS---GASLYIYPAGELKNLPFGGIWDATLKL-RVKRRY 183
Db 132 DACYTSV-----INMNAACQGRSLKRLPSEELAKIPTSGTWKATLVLDLWGG 183
QY 184 SETYGTYTINIKLTD--KGNIQIWLQF--KSDARVDNLNRPTGGTYIGNSVDMCFY 240
Db 184 DPLGTSTTIDILNVDHPAENAIYFPQGTATPRVDNLNRHMASQMSGRANLDMCLY 243
QY 241 DGYSTNSSLEIRFQDNNPKSDGKFLYLRKINDTKEIATYLSLLAGKSLTPPTNGTSLNI 300
Db 244 DG-GVKARSLOWKIEGSKSGTGFQVKSADSAT--IDYAVSMYGGRSIPVTRGVFEFL 300
QY 301 ADAASLETNNRITAVTMPEISVPLVCPGRQLQDAK---VENPAGQWGMNINVTFTPS 357
Db 301 DNVDKAATR-----PVLPGRQAVRCVPLTLTTPPNIREKSGEYQGITLVMLMG 355
QY 358 SQT 360
Db 356 TQT 358

RESULT 8
Q83JU1 PRELIMINARY; PRT; 245 AA.
AC Q83JU1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Orf, partial conserved hypothetical protein.
GN SP3134.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=301 / Serotype 2a;
RC MEDLINE=2272436; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng E., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RA "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015325; AAN44606.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 245 AA; 27602 MW; 4D27007AE2F0B1A3 CRC64;
Query Match 14.4%; Score 272.5; DB 16; Length 245;
Best Local Similarity 32.0%; Pred. No. 8.1e-13;
Matches 73; Conservative 37; Mismatches 89; Indels 29; Gaps 9;

QY 153 LYLILPAGELKNLPFGGIWDATLKLVRKRYSTGYTYITIKLTKDGNIGIWLQPF- 211
Db 27 LTYLQTEKKNLPIGGVW---KGVKLNHSNPAQDYFANITLNTLDPHVIDVFPEFA 82
QY 212 KSDARVDNLNRPTG---GGTYI-GRNSVDMCFYDGYSTNSSSLEIRFQDNN----PKSDG 263
Db 83 HATPRVQLDHLPTGVSNGSNYAQDLTMDLWCLYDGFNGNAISYEIMLKDEGRPAAGRDG 142

```

```

QY 264 KYLRK---INDTKETAYTUSLLLAGKSLTPTNGTSLNIADAASL---ETNWRITAV 316
DB 143 YFSIYRQGGTTDEGERIDYR-----KYNPFTGGQIDVRNENMVWNSINLKRVRP 196
QY 317 TMEPISVPVLCWPGRLQDA---KVENPEAGYMGNIINVTPESSQTL 361
DB 197 VLPGIRYAVMCPVPTLTLAVDRFSVMDKQGYMGKLSVIFTPSLPTI 244

RESULT 9
QYUBH8 PRELIMINARY; PRT; 210 AA.
AC QYUBH8;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN S3341
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=2457T; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; A2016989; AAP18419.1; --
KW Hypothetical protein.
SQ SEQUENCE 210 AA; 23516 MW; BBD109DC114A2486 CRC64;

Query Match 13.6%; Score 256.5; DB 16; Length 210;
Best Local Similarity 31.5%; Pred. No. 1.1e-11;
Matches 69; Conservative 36; Mismatches 85; Indels 29; Gaps 9;

QY 162 LKNLPGGIWDATLKLVRKRRSETGTYYTINIKLTKDGN-QIWLQPF-KSDASVDLN 220
DB 1 MKNLPIGGVW---KGKVLHNSPQAQYFANITLNTLDPNH-DVFPFPAHATPQVQLD 56
QY 221 LRPTG---GTYI-GRNSVDMCFYDGYNSSSLSIRFQDN---PKSDGKFLRK--- 269
DB 57 LRPTGVSNSVAQDLTMDCLYDGFNGNAISYRMLKDEGRPAAGRDGYFSYRGG 116
QY 270 -INDTKETAYTUSLLLAGKSLTPTNGTSLNIADAASL---ETNWRITAVTMEPISVPV 325
DB 117 TTTDEGERIDYR-----KYNPFTGGQIDVRNENMVWNSINLKRVRPVLPFGIRYAV 170
QY 326 LCWPGRLQDA---KVENPEAGYMGNIINVTPESSQTL 361
DB 171 MCVPTLTLAVDRFSVMDKQGYMGKLSVIFTPSLPTI 209

RESULT 10
QYUBH8 PRELIMINARY; PRT; 1801 AA.
AC QYUBH8;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hemolysin.
GN FN0291.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Barton A., Gardner W., Grechkin G., Zhu J.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Ponstein M., Kyridides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; A2010541; AAL94497.1; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008638; Haemaggl.
DR InterPro; IPR000508; Peptidase_S26.
DR InterPro; IPR001452; SH3.
DR Pfam; P705860; Haemaggl_act; 1.
DR Pfam; P300018; SH3; 1.
DR PROSITE; PS00761; SPASE_1_3; 1.
KW Complete proteome.
SQ SEQUENCE 1881 AA; 204375 MW; D635156A4EFA4877 CRC64;

Query Watch 6.7%; Score 127; DB 16; Length 1881;
Best Local Similarity 23.1%; Pred. No. 1.2; Indels 122; Gaps 19;
Matches 96; Conservative 45; Mismatches 122; Gaps 19;

QY 1 MKNLFIPTLFSSVLTFAV-----SADKIPGDESITNFGPRDRNESSPKHN- 49
DB 5 LKLLTAIFMLFHLIISLADGIVPDNSASKNLQVKAANGVPLVNIAP-DNNGTS--HNV 61
QY 50 -----LNNHITAYSESHILYDRMTFLCLSSHNTLNCACETSPSSSSVSG-- 96
DB 62 YKDYNVDRGSAFLNN-----SKLTNSQLGGLI 89
QY 97 ETNITLQTEKPSLIKREL-----CIKGYKOLLFKSVNC-----PSGLTLNGAHF--NCN 144
DB 90 YGNPLQNSKEASTINEVSGVNGKRIEIGQIVGKRYANILANPNIINGACFINTGN 149
QY 145 KNAASASLYLIP-----AGEIKNLPFGGIWDATLKLVRKRRSETGTYYTINIK 197
DB 150 VFTTIGRGNLNPCKMTEIAG--KGLDLRNINKAEILARVAELSAPIYGGVEVNLKG 207
QY 198 LNDKIGNIQWLQFSDARVDNLNLRPTGGTYGRNSVDMCFYDGYSTSSSLRPO-- 255
DB 208 NQKSN---KPEYALDAR-----ALGSIYAGRINI-----IVNEDGVGVKTQAP 248
QY 256 -----DNPKSDGKFLR-----KINDDTKEIAYTUSLLLAGKSLTPTNGTSLNIA 301
DB 249 MYATKGVVISKGVKYLQTKAKRIKSSITETIG---SKLLAENAIKSKT--SN-- 303
QY 302 DAASLETNWRIT 314
DB 304 -SQQIRAN--NIT 314

RESULT 11
QYUBH8 PRELIMINARY; PRT; 700 AA.
AC QYUBH8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative polymorphic membrane protein (Fragment).
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=POS;
RA Laroucau K., Souriau A., Rodolakis A.;
RT "Isolation of a new pmp sequence and evidence of pmp polymorphism in
RT serotype-1 Chlamydia psittaci strains."

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF243415; AAL36959.1; --
 DR InterPro; IPR003368; Chlamydia_PMP.
 DR Pfam; PF02415; Chlamydia_PMP; 1.
 DR TIGRFAMs; TIGR01376; POMP_repeat; 3.
 FT NON_TER 700 700
 SQ SEQUENCE 700 AA; 74200 MW; 2C38E504AF48D1D1 CRC64;

Query Match 6.1%; Score 115.5; DB 2; Length 700;
 Best Local Similarity 18.3%; Pred. No. 2.5; Indels 135; Gaps 18;
 Matches 73; Conservative 63; Mismatches 124; Indels 135; Gaps 18;
 QY 42 NESSPKHILNNHITA-----YSSSHLYDRMTFLCL---SHNTLNGAC-----53
 DB 27 NEILTSSDSYNGNVSDPEVXETTSAGIATCEGNVCISYAGKOSPLNKSCFSETTENLS 86
 QY 84 -----PTSENPSSSSVSGETNNITLQFTKESLKLKRELOIKGYKQLLPKSVN 129
 DB 87 FIGNGYTLFCFNDNITTTASNGAINVSGD-----QKTLNVSGFS--LFSCAH 130
 QY 130 CPSGLT-----LNSAHF--NCNK-----NAASGASLYLYIP-----A 159
 DB 131 CPPTGTGYGAIOATKGVTSTGNNKLIFDNNCSGEGGAIKCATGSAELKLENSVWFS 190
 QY 160 GELKMLPFGGIMDAITLKLVRKRYSEYTYGTITINIKLTDKGNIQIWLPOPKSDARVDL 219
 DB 191 GNSSQKKGAIY--TKLITADGPTLPSNNVSAS---SPKGG-AICLDTSSECSLTA 244
 QY 220 NL-----RPTGGGTIVIGNVDM-----CFYDGYSTNSSSLR 251
 DB 245 NLGDITFDGNKVKIKINGSGSTVRNAIDLGSGGKFTKLNAAKEGFGIFPHDPIANTGGSTE 304
 QY 252 IRFDNNPKSD---GK--FYLRKINDDTKEIAYTL-----SILLA 286
 DB 305 IEL--NATESDITTYTKIVFSGSKLDEEKTVPDNLKSYFKQPLKICAGSLVLKDGVTLE 362
 QY 287 GKSLTPTNGTSLNIADAASLETNNRITAVTMTPEISVPV 325
 DB 363 AKKITQKSTVVDLGTITQTSSSGETITLNLINI 401

RESULT 12

P71132
 ID P71132 PRELIMINARY; PRT; 847 AA.
 AC P71132; 1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE POMP91A.
 OS Chlamydia abortus.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydiaophila.
 OX NCBI_TaxID=83555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S26/3;
 RX MEDLINE=96406378; PubMed=6810511;
 RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
 RT "Identification of a multi-gene family coding for the 90 kDa proteins
 of the ovine abortion subtype of Chlamydia psittaci";
 RT FEMS Microbiol. Lett. 142:277-281 (1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S26/3;
 RX MEDLINE=98187897; PubMed=9529048;
 RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
 RT "Molecular cloning and characterization of the genes coding for the
 highly immunogenic cluster of 90-kilodalton envelope proteins from the
 Chlamydia psittaci subtype that causes abortion in sheep";
 RT Infect. Immun. 66:1317-1324 (1998).
 RL EMBL; U65942; AAC15921.1; --
 DR InterPro; IPR006315; Autotransport.
 DR InterPro; IPR003368; Chlamydia_PMP.
 DR Pfam; PF02415; Chlamydia_PMP; 1.
 DR

DR TIGRFAMs; TIGR01414; autotrans_bar1; 1.
 DR TIGRFAMs; TIGR01376; POMP_repeat; 3.
 SQ SEQUENCE 847 AA; 90694 MW; 754C958E7F11179E CRC64;
 Query Match 6.1%; Score 115.5; DB 2; Length 847;
 Best Local Similarity 21.6%; Pred. No. 3.2; Indels 131; Gaps 21;
 Matches 94; Conservative 58; Mismatches 153; Indels 131; Gaps 21;
 QY 10 LFFSVLFTFAVSADKIFGDSITNIFQPRDRNESPSPKHILNNHITAYSSHTLYDRMT 69
 DB 9 LSSSLIVSNLSYSE--PDQKTLTSAHGVNGTNSNP-----FNPLSTSNNGTIYCTG 62
 QY 70 FCLSL-----SINTLNGACPT-----SENPSSSSVSGETNITL 102
 DB 63 NICIAYAGLDGSGJSSSCFTDTAGNLFLGNGYTLFCFONITTOSSHHPGAISSVG-TNKT- 120
 QY 103 QFTKRSLLIKRELQIKGYKQLLPKSWCPSPGLTNSA-----HFNCKN 145
 DB 121 -----LDISGFS--LFSYAYCPPGATGYGAIKAVGNITIKONSSLVFHKNCST 166
 QY 145 NAASGASLYLYIPAGELK-----NLPP-----GGTWDATLKLVRKRYSEYTYTTL 192
 DB 167 GEGGAIQCKKASSSEAELEKIENNQLVFAENSSSSGGAIYAD-KLTISSGGPTLFSNNSV 225
 QY 193 NIT-----IKLTDKGNIQIWLPOPKSDARVDL--LRPTGGGTIVIGNVDM----- 237
 DB 225 SASSPKGGAICIKDSGG--ECSLTADLGDTTFDKKLIKINGSGSPVTRNSIDLSSGKFT 284
 QY 238 -----CFYDGYV--TNSSSLRIFQDNNPKSDGK--FYLRKINDDTKEIAYTL----- 281
 DB 285 KLNAAKEGFIPIYDPIITGGSGSELINIKQTDVDT--GKIVFSGERLSDEKKVAANLKSD 343
 QY 282 -----SILLAG-----KSLTPTNGTSLNIADAASLET--NWNRITAVTMTPEISV 323
 DB 344 FKQPLKIGSGSLIKDGVTLTKSTQTQEGATVMDLGTTLQTPSSGGETITLNLINIV 403
 QY 324 PVLCPGRQLDAKVE 339
 DB 404 ASLGGGVAPDPAKVE 419

RESULT 13

Q8Q7H7
 ID Q8Q7H7 PRELIMINARY; PRT; 865 AA.
 AC Q8Q7H7;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN ENV.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=99CMA304;
 RX MEDLINE=21849375; PubMed=11830674;
 RA Yamaguchi J., Vallari A.S., Swanson P., Rodelle P., Kaptue L.,
 RA Nansop C., Zekeng L., Gurtler L.G., Devare S.G., Brennan C.A.;
 RT "Evaluation of HIV Type 1 Group O isolates: Identification of five
 phylogenetic clusters";
 RT AIDS Res. Hum. Retroviruses 18:269-282 (2002).
 RL EMBL; AF383244; AAL98866.1; --
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env_Gr41.
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 DR ALDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
 NON_TER 1 1

SQ SEQUENCE 865 AA; 97118 MW; 996883862AA31ACA CRC64;
Query Match 6.0%; Score 114; DB 15; Length 865;
Best Local Similarity 21.5%; Pred. No. 4.2;
Matches 93; Conservative 56; Mismatches 150; Indels 134; Gaps 22;
QY 21 VSADKTPGDSITNFGPRDRNESSPKHNLNNH-----ITAYSSEHLYDRMTFL 71
DB 75 VPTDTPFYPLHN-----TDKFIWENVMYKQMGDIIDWEOSLPCVKMTFL 125
QY 72 C-SSNMT-NGACPTSENSSSVS---GETNITLQF-EKSLIKRELQIKYKQLLEK 126
DB 126 CVMQNTSLN---DTKSNFTSSPENLMKCEFNVTWVKQKKAQALFVSDMLKLDK 182
QY 127 SYNCPSGLTNGAHFNCKNAASGA-----SYLYIPAG-----ELKNPFGG--- 169
DB 183 NITNTMTYLT---NONSSTISOACPKVSFEPIPIHYCAPAGAYAKCNIEFNSTGTC 238
QY 170 -----INWATL---KLRV-KRYSETYG-TYTNITIKLTDK--G 202
DB 235 KNTTVMCTHGKPTYSTOLLNGTLGKIRIMTKNISDNVNIIVTLASTLKTCRPG 298
QY 203 NIQIWLPOKSDARVDNLNRPCTGGTYIGRNSVDMCFYDGYSTNSGSLRIFODNNPKSD 352
DB 299 NMTVQOVSTGPNWAWYMDLR-----GTERNTSRVAYCEY-----NSTDWE-----RTLKQT 344
QY 263 GKPYLRKINDTKEIAYTLLSLLAG-----KSL-----TPTNGTSL 298
DB 345 AERFLELVN-NTKKVDMTFNSGGDPEVANLHFNCHGEFFYCNSTSLFNFTSCNGSTC 403
QY 299 NTAADAASLETNWRI-----TATMPEISVPVLQWPGRLDLDKAVNPEA 343
DB 404 NVTKSN-NSTNTRIPCLRQVVKSWTQGGSLGYAPPFIRGNLTOMENITGLMQLMDQP-- 460
QY 344 GQYMGNIWTFP 356
DB 461 ---WNHNSATFRP 470
RESULT 14
Q81521 PRELIMINARY; PRT; 2646 AA.
AC Q81521
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Erythrocyte membrane protein 1 (pEMP1).
GN PF11950W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7; PubMed=12358864;
RX MEDLINE=22255705; PubMed=12358864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Rene V., Shalim S.J., Suh B., Peterson J., Angiucci S.,
RA Perica M., Allen J., Selegut J., Haft D., Mather W.W., Vaidya A.B.,
RA Martin D.N.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
ET "Genome sequence of the human malaria parasite Plasmodium falciparum.";
RL Nature 415:498-511(2002).
DR EMBL; AE014850; AAN36476.1; ..
DR GO; GO:0005539; F:Glycosaminoglycan binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR004258; PfEMP.
DR Pfam; PF03011; PfEMP; 2.
SQ SEQUENCE 2646 AA; 300284 MW; C91D257F3DEF8717 CRC64;

Query Match 6.0%; Score 113; DB 5; Length 2646;
Best Local Similarity 24.8%; Pred. No. 21;
Matches 77; Conservative 37; Mismatches 128; Indels 68; Gaps 15;
QY 23 GDESITNFGPRDRNESSPKHNLNNHITAYSESHTLYDRMTFLCLSSNTLNGACPTG- 86
DB 695 GDS--TAIQIGIKTKLEKKKKGAN-----ISEEKTILDEFNLHETKDAETCKNCEPRKF 747
QY 87 ENPSSSVSGEINILQFTEKESLIKRELQIKYKQLLEKSYNCPGSLTNSAHFNCKN 146
DB 743 KNPCSGDTSN--KQVEAVANTVAQLQGAQKOL-----HNGSRN 789
QY 147 AASGASLYIIPAGELKNLPFGGICWDATLKLKRVRYSETYTYTITIKLTDKNI-- 204
DB 790 ALKGNLQNAKINNGKPN-----PLTDACQITKXHSNGK-----DSNNPCNKNRLLK 838
QY 205 --QIWLPOKSD-ARYDL-----NLRPTGGTYIGRNSVDMCFYD--YSTNS 247
DB 839 IQQVW--SIKNDTSYTDVMPRPQRQHVCTSNLEKLYASVIGSNVNDKFLVEVLHAKS 896
QY 248 SSLEIRFQNNPKSDGKFIYRKINDT-KEIAYT---LSLLAGKSLTPTNGTSLNADA 303
DB 897 EAEFIKKYNEKQNDCKNGLRKQDATTCAIRYSFADIGDIINGKOLWDDNN-----DA 950
QY 304 ASLETNWNEI 313
DB 951 KSLQTNLKA 960
RESULT 15
Q8XNW2 PRELIMINARY; PRT; 743 AA.
AC Q8XNW2
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein CP0220.
GN CP0220C.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
ET "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79526.1; ..
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006635; NEA_transpt.
DR SMART; SM00725; NEAT; 4.
DR TIGRfam; TIGR01167; LPXTC_anchor; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 743 AA; 81149 MW; A7C451E39C5A5E545 CRC64;
Query Match 5.9%; Score 111.5; DB 16; Length 743;
Best Local Similarity 18.8%; Pred. No. 5.4;
Matches 64; Conservative 66; Mismatches 115; Indels 95; Gaps 13;
QY 7 IFTLFFSSVLTFAVSADKTPGDSITNFGPRDRNES-----SPKHNI-LNNHITAYSES 61
DB 88 LMTVVFNSSLYGFNMNIEVSGGEGALKIEBKDKSITFEVPSPDCKVKIGLFTMMGRK 147
QY 62 HTLYDRMTFLCLSSHNTLN--GACPTSENSSSV-----SG-----ETNITLQ 103
DB 148 VELF-----LVNDMTVNLDEAPTNNAKDISVTQGDADILLGSGVGTGKSDSNLKVE 201
QY 104 PTEKRSLIK-RELQIKYKQLLEKSYNCPGSLTNSAHFNCKNKAASGASLYIYPAGEL 162

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:11:46 ; Search time 18 Seconds
(without alignments)
1044.296 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFIFLTFSSVLTFFA.....EAGQTMGNINVTFTSSQVL 361

Scoring table: ELOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558.5	82.6	360	CFAB_ECOLI	P25734 escherichia
2	120	6.4	2650	YEEU_YEAST	Q8x8v7 escherichia
3	114	6.0	650	SAGI_YEAST	P20840 saccharomyc
4	109	5.8	2358	YEEU_ECOLI	P76347 escherichia
5	108	5.7	524	GASJ_YEAST	Q03655 saccharomyc
6	107	5.7	304	YDEQ_ECOLI	P77588 escherichia
7	100	5.3	737	ALYS_ENTFA	P37710 enterococcu
8	99	5.2	802	SYFB_LIGIN	Q92016 listeria in
9	99	5.2	1849	IGA4_HABIN	P45386 haemophilus
10	98.5	5.2	608	YD93_METJA	Q58788 metnanococ
11	98	5.2	758	SP21_YEAST	P35209 saccharomyc
12	98	5.2	1157	N133_YEAST	P36161 saccharomyc
13	95.5	5.1	713	CDG2_PAPMA	P31833 paenibacill
14	95	5.0	413	RPLM_YEAST	P30775 saccharomyc
15	95	5.0	2334	WAPA_BAGSU	Q07833 bacillus su
16	94.5	5.0	959	N100_YEAST	Q02629 saccharomyc
17	94	5.0	471	LEU2_BUCRP	P48573 buchnera ap
18	93.5	5.0	412	THEG_SHEEP	P50450 ovis aries
19	93	4.9	1167	CLAA_BACTU	P56956 bacillus th
20	93	4.9	1773	DIP2_DROME	Q9w089 drosophila c
21	92.5	4.9	320	CYF_CYACA	Q9c1s4 cyanidium c
22	92.5	4.9	544	AGM1_CANAL	Q9p4v2 candida alb
23	92.5	4.9	1061	OAR_MYXAA	P38370 myxococcus
24	92.5	4.9	2278	PAB1_YEAST	P34756 saccharomyc
25	91.5	4.9	500	SVK_EUCBP	Q89acs buchnera ap
26	91.5	4.9	869	CFAC_ECOLI	P25733 escherichia
27	91.5	4.9	1783	Y468_MYCCE	Q49460 mycoplasma
28	91	4.8	802	SYFB_LISNO	Q9y7q1 listeria mo
29	90.5	4.8	309	ELTB_CLOPE	P01558 clostridium
30	90.5	4.8	863	MCN4_XENLA	P30664 xenopus lae
31	90.5	4.8	1019	ENTK_HUMAN	P88073 homo sapien
32	90.5	4.8	1328	HUS2_SCHPO	Q09811 schizosacch
33	90.5	4.8	2193	POLG_HE71M	Q66479 h. genome po

RESULT 1
ID CFAB_ECOLI STANDARD; PRT; 360 AA.
AC P25734;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE CFA/I fimbrial subunit B (Colonization factor antigen I subunit B).
GN CFAE.
OS Escherichia coli.
OG Plasmid NTP513.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
[1]
SEQUENCE FROM N.A.
RC STRAIN=Enterotoxigenic;
RX MEDLINE=89330163; PubMed=2569152;
RA Hamers A.M., Pel H.J., Willshaw G.A., Kusters J.G.,
van der Zeijst B.A.M., Gastra W.,
"The nucleotide sequence of the first two genes of the CFA/I fimbrial
operon of human enterotoxigenic Escherichia coli.";
RL Microb. Pathog. 6:297-309(1989).
[2]
SEQUENCE FROM N.A.
RP MEDLINE=92329981; PubMed=1352712;
RX Jordi B.A.M., Willshaw G.A., van der Zeijst B.A.M., Gastra W.,
"The complete nucleotide sequence of region I of the CFA/I fimbrial
operon of human enterotoxigenic Escherichia coli.";
RL DNA Seq. 2:257-263(1992).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMEL; N55661; AAC41417.1; --
DR PIR; D56617; D56617.
DR InterPro; IPR008957; FN III-like.
KW Antigen; Fimbria; Plasmid.
SQ SEQUENCE 360 AA; 39903 MW; 691509B63A8E69CE CRC64;

Query Match 82.6%; Score 1558.5; DB 1; Length 360;
Best Loca. Similarity 81.2%; Pred. No. 3.6e-111;
Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;
Qy 1 MNKILFIFLTFSSVLTFFA...EAGQTMGNINVTFTSSQVLNINVTAYSE 60
Db 1 MNKILFIFLTFSSVLTFFA...EAGQTMGNINVTFTSSQVLNINVTAYSE 60
Qy 61 SHLYDMSFLCLSSQNTLNACAPSSDPATIDGTNITLQFTKRSIKRELQIKGY 120
Db 61 SHLYDMSFLCLSSQNTLNACAPSSDPATIDGTNITLQFTKRSIKRELQIKGY 120

```

QY 121 KQLPKSVNCPGLTNSAHFNCKNAAGSASLYIPAGELKNLPFGGLWDATLKLRYK 180
DB 121 KQLPKSVNCPGLTNSAHFNCKNAAGSASLYIPAGELKNLPFGGLWDATLKLRYK 180
QY 181 RRYSEYGVYVITINIKLTDKGNQIWLPOFKSDARVDLNLRTGGTYIGRNSVDMCFY 240
DB 181 RRYSEYGVYVITINIKLTDKGNQIWLPOFKSDARVDLNLRTGGTYIGRNSVDMCFY 240
QY 241 DGSTNSSSLLEIRPDNNPKSGDKFYLRKINDTKELIATLSLLIACKSITPTNGTSLMI 300
DB 241 DGSTNSSSLLEIRPDNNPKSGDKFYLRKINDTKELIATLSLLIACKSITPTNGTSLMI 300
QY 301 ADAASLETNWRITAVTNPISVPLVCPGRLQDQKVENPEBAGQYMGINIVFTPTSSQT 360
DB 301 -NTASLETNWRITAVTNPISVPLVCPGRLQDQKVENPEBAGQYMGINIVFTPTSSQT 360
QY 361 L 361
DB 360 L 360

RESULT 2
YEEU ECO57 STANDARD; PRT; 2660 AA.
AC Q8X8V7; O8X2B9; O8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein yeeJ.
GN 23135 OR EGS2775/EGS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83324;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11256551;
RA Perna N.T., Plunkett G., Burtand V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RI "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=2115231; PubMed=11256796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by ard for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005423; AAG57041.1; -
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.

```

```

DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big-1.
DR InterPro; IPR003355; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR InterPro; IPR000601; PKD.
DR Pfam; PF02369; Big-1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00089; PKD; 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 929
FT DOMAIN 931 1033
FT DOMAIN 1042 1132
FT DOMAIN 1134 1236
FT DOMAIN 1245 1335
FT DOMAIN 1337 1439
FT DOMAIN 1448 1539
FT DOMAIN 1548 1652
FT DOMAIN 1653 1750
FT DOMAIN 1751 1855
FT DOMAIN 1856 1957
FT DOMAIN 1963 2056
FT DOMAIN 2065 2156
FT DOMAIN 2157 2252
FT DOMAIN 2254 2355
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 6.4%; Score 120; DB 1; Length 2660;
Best Local Similarity 20.4%; Pred. No. 0.79; Mismatches 156; Indels 90; Gaps 15;
Matches 80; Conservative 66;

QY 13 SSVLFTFAVSADKIPGDESIINIFGPRDRNESSPKENILNNHITAY---SESLTLDKMT 69
DB 737 SAKIALLSASNNGLANENAA--VSVVADEGS---NPINDHTVTPAVLSSGSAFENNQN 793
QY 70 FLCLSHNTINGACPTSENPSSSVSGEITNIQTFTKSLI-----K 112
DB 794 ----TAKTDVNGLA--TFDLKSKQEDNTVEVLENGVKQTLIVSVGDSSTAQVDLQSK 848
QY 113 REIQIKGY-KQLLFKSVNCPGSLTNSAHFNCKNAAGSASLYIPAGELKNLPFGGIW 171
DB 849 NEVADGNSATWATVDRDAKGNLNDVKVTNVNSAAAKLSQTEVNSHD-----GIA 901
QY 172 DATLKLVRKRYSEYGVYVITINIKLTDKGNQIWLPOFKSDARVDLNLRTGGTYIG 231
DB 902 TATLT-----SLKNGDYTVTASVSSGSAQNVIFIGQSTAALPLSV-PSGDITV-- 951
QY 232 RNSVDMCFYGVYVITINSSSLLE---IRPDNNPKSGDKFYLRKINDTKELIATLSLLIAG 287
DB 952 -----TNTAPLEMTATLQDKNGNPLKDKBITFVSFVND-----VASRFSIENSG 994
QY 288 KSLTPINGTSL-----NIADAASLETNWRITAVTNPISVPLVCP 327
DB 995 KGMVDSNGTATIASLTGLTLAGHMTATRLANSVSDTQPMTFVADKDEAVVVLCTSKAEII 1054
QY 328 WPG--RLQLDAKVENPEBAGQYMGINIVFTPTPS 357
DB 1055 GNGVDETTLTATVKDP-PDNVKNLSVVRFTS 1085

RESULT 3
SAG1 YEAST
ID SAG1 YEAST STANDARD; PRT; 650 AA.
AC P20840.
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Alpha-agglutinin precursor (AG-alpha-1).
DE SAG1 OR AGAL1 OR YJR004C OR J1418.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```


RESULT 4

YEEJ ECOLI
 ID YEEJ ECOLI STANDARD; PRT; 2358 AA.
 AC P76347; P94750;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein yeeJ.
 GN YEEJ OR B1978.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97251358; PubMed=9278503;
 RA Blattner F.A.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
 RA Riley M.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;
 RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
 RA Mau B.; Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 279:1453-1474 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251358; PubMed=9087040;
 RA Itoh T.; Aiba H.; Baba T.; Fujita K.; Hayashi K.; Inada T.;
 RA Isono K.; Kasai H.; Kimura S.; Kitakawa M.; Kitagawa M.;
 RA Makino K.; Miki T.; Mizoguchi K.; Mori H.; Mori T.; Motomura K.;
 RA Nakade S.; Nakamura Y.; Nishimoto H.; Nishio Y.; Oshima T.;
 RA Saito N.; Sempel G.; Seki Y.; Sivasubraman S.; Tagami H.;
 RA Takeda J.; Takemoto K.; Wada C.; Yamamoto Y.; Horiuchi T.;
 RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392 (1996).
 CC -!- SIMILARITY: Contains 13 Big-1 domains.
 CC -!- SIMILARITY: Belongs to the intimin/invasin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE000289; AAC75042.1; ALT_INIT.
 DR EMBL; D90837; BAA15800.1; ...
 DR EMBL; D90836; BAA15799.1; ALT_INIT.
 DR EcoGene; EG13378; yeeJ.
 DR InterPro; IPR003344; Big_1.
 DR InterPro; IPR003535; Intimin.
 DR InterPro; IPR008964; Invasin_intimin.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR000601; PKD.
 DR Pfam; PF02369; Big_1; 13.
 DR PRINTS; PR01369; INTIMIN.
 DR SMART; SM00634; BID_1; 13.
 DR SMART; SM00257; LysM; 1.
 DR SMART; SM0089; PKD; 6.
 KW Hypothetical protein; Repeat; Complete proteome.
 FT DOMAIN 738 834
 FT DOMAIN 834 834
 FT DOMAIN 840 931
 FT DOMAIN 932 1033
 FT DOMAIN 1042 1137
 FT DOMAIN 1146 1237
 FT DOMAIN 1246 1350
 FT DOMAIN 1351 1448
 FT DOMAIN 1449 1553
 FT DOMAIN 1554 1655

FT DOMAIN 1661 1754
 FT DOMAIN 1763 1853
 FT DOMAIN 1855 1950
 FT DOMAIN 1952 2053
 FT DOMAIN 2055 2150
 FT COMPLECT 105 105
 FT COMPLECT S -> G (IN REF. 2).
 SQ SEQUENCE 2358 AA; 248559 MW; 2322497503P631ED CRC64;
 Query Match 5.8%; Score 109; DB 1; Length 2358;
 Best Local Similarity 20.0%; Pred. No. 4.6;
 Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;
 QY 13 SSVLFFFAVSADKIPGDESITNIFGPRDNESSPGHNLNKHITAY---SESHTLYDRMT 69
 DB 737 SAKIATLSASNNGLANVAANTVSVNVADEGS---NPINDTVTFAVLSGSATSFNQCN 793
 QY 70 FLCLSSHNILNACPTSENPSVSSVSGETNITLQPTKRSLSI-----K 112
 DB 794 ----TAKTDVNGLA-TFDLKSSKQEDNTVEVLENGVKQTLIVSFWGDSSTAQVDLQSKS 848
 QY 113 RELQIKGYKQL-LFKSVNCPGSLTNSAHFNCKNAASGASLYLIPAGELKNLPFGGIW 171
 DB 849 NEVVADGNDSTMTATVDEAKGNLLNDVMVTFNVAABAKLSQTEVNGHD-----GTA 901
 QY 172 DATLKRVKRYSEYTYTYTITIKLTDKGNIQIWLPOFKSDARVDNLNLRPTGGTYIG 231
 DB 902 TATLT-----SLXNGDYRVTA SVSSGSQANQQVNFIDQQTALTLVSV-PSGDIIV-- 351
 QY 232 RNSVDMCFYDGYSTNSS---SLETRFQDNNEPKSDGKFLRKINDDTKEIAYTLISLLAG 287
 DB 952 -----TNTAPQYMTATQDKNGNPLKDEITFSVPND-----VASKFSISNGG 394
 QY 283 KSLTPTNGTSL-----NIADAASLETWNRAITATVMEISVPLVC 327
 DB 995 KGMTDSNGVAIASLTCTLAGTETIMARLANSVSDAQPMTFVADKDRVAVVLTQSKAEII 1054
 QY 328 WPG--RLQIDAKVENPEACQYMGNIIVTPT 355
 DB 1055 GNGVDETTLTATVKDP-SNHPVAGITVNT 1083
 RESULT 5
 GAS3_YEAST STANDARD; PRT; 524 AA.
 ID GAS3_YEAST
 AC Q03655;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Glycolipid anchored surface protein 3 precursor.
 GN GAS3 OR YMR215W OR YMR261.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=9713268; PubMed=9169872;
 RA Bowman S.; Churcher C.M.; Badcock K.; Brown D.; Chillingworth T.;
 RA Connor R.; Dedman K.; Devlin K.; Gentles S.; Hamlin N.; Hunt S.;
 RA Jørgensen K.; Lyle G.; Moule S.; O'Connell C.; Pearson D.; Rajandream M.A.;
 RA Rice P.; Skelton J.; Walsh S.; Whitehead S.; Barrell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 XIII.";
 RL Nature 387:90-93 (1997).
 RN [2]
 RP IDENTIFICATION.
 RX MEDLINE=20529944; PubMed=11079560;
 RA Pardo M.; Ward M.; Bairs S.; Molina M.; Blackstock W.; Gil C.;
 RA Nombela C.;
 RT "A proteomic approach for the study of Saccharomyces cerevisiae cell
 wall biogenesis.";
 RL Electrophoresis 21:3396-3410 (2000).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor

```

SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12.";
Science 277:1453-1474(1997).
[2]
SEQUENCE FROM K.A.
STRAIN=K12.
MEDLINE=97251357; PubMed=9097039;
Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
Kasai H., Kashimoto K., Kimura S., Kiteakawa M., Kitagawa M.,
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Makade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
Sampai G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
Takenoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
"A 570-Kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
DNA Res. 3:363-377(1996).
-|- SIMILARITY: TO TYPE-1 FIMBRIAL SUBUNITS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB000247; AAC74575.1; -.
EMBL; D90792; BAA15175.1; -.
EMBL; D90793; BAA15183.1; -.
PIR; A64904; A64904.
DroGene; EG3739; ydeQ.
InterPro; IPR038966; Adhes.bact.
InterPro; IPR000259; Fimbria.
Pfam; PF00419; Fimbrial; 1.
Hypothetical protein; Fimbria; Signal; Complete proteome.
SIGNAL
27 304 POTENTIAL.
CSAIN 1 26
SEQUENCE3 304 AA; 32069 MW; 8153C86E3087D99A CRC64;
Query Match 5.7%; Score 107; DB 1; Length 304;
Best Local Similarity 21.0%; Pred. No. 0.48;
Matches 75; Conservative 51; Mismatches 125; Indels 106; Gaps 19
QY 3 KILP-IPTLPFSSVLFTFPAVSADKIPGDESI-----TNIFGRDRNESSPKINI---LNKH 54
DB 8 KVLFGIYLLLMAGVFAFSCNVD---GGSSIFAGTTSVYNLD-PVIFGQNLVVDLSQH 63
QY 55 ITAYSESHTYDRMTPLCLSSH-NILNGACPTSENPSSSVSGETNITLQTFEKSLIKR 113
DB 64 ISCVNDYGGWVD-----TDHNLVQG-----SAPAG----- 89
QY 114 ELQTKGYK-QLLFKSVNCPSGILTNLSAHFNCKNAASGASIVLYIPAGELKNLPFGGIWD 172
DB 50 --SLSQYKSGSLYNNVVIYPPELLTNTNVLIDGDKTPMPLPLKLYI-----TPVGAAGG 140
QY 173 ATLK-----LRVKERYSETYGT-----YTINITIK-----LTKGNIQIWLPOFK 212
DB 141 VVIRAGEVIAIRHMYKIALTSGNPRFTWAIISNNVVMPTGGCTVDSRNVVDLPDPF 200
QY 213 SDARVDNLIRPTGGGTIGRNSVDMCFYDGYSTNSSLSLEIRPDNNPKSDGFYLRKIND 272
DB 201 GSAEIPL-----GVYCS-SEQKLISFLSGATDSSRQV-FANTAP-----D 239
QY 273 DTKRIAYTLISLLACK-----SLPTNGTSLNIADAASLETNWRITAVTMEI 321
274 NTYASGVCTVWNGKILATGENTSVIGTVNKSQVPLGHSATYGTGCKNKGAGIVQSV 296

```

```

RESULT 7
ALYS_ENTFA
ID ALYS_ENTFA STANDARD; PRT; 737 AA.
AC P37710;
DT 01-OCT-1994 (Rel. 30, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Autolysin precursor (EC 3.2.1.1-) (Peptidoglycan hydrolase)
DE (Aer-a-glycosidase).
GN BF0799.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91358349; PubMed=1679432;
RA Bellevue C., Polvin C., Trudel J., Asselin A., Bellemare G.;
RT "Cloning, sequencing, and expression in Escherichia coli of a
RT Streptococcus faecalis autolysin.";
RL J. Bacteriol. 173:5619-5623(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=V583 / ATCC 700602;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Barerjei L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Unayam L., Brinkac L., Beanan M.,
RA Tagherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Opatkin J., Hansen T., Shetty C., Khouri H.,
RA Ueberback T., Radune D., Ketchum K.A., Dougherty S.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
CC -!- FUNCTION: Hydrolyzes the cell wall of E.faecalis and
CC M.lysodisaccharus. May play an important role in cell wall growth
CC and cell separation.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- DOMAIN: LysM repeats are thought to be involved in peptidoglycan
CC binding.
CC -!- SIMILARITY: Belongs to family 73 of glycosyl hydrolases.
CC -!- SIMILARITY: Contains 6 LysM repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EXBL; M58002; AAA67325.1; -.
CC EXBL; AE016949; AAC08013.1; -.
CC PIR; A38109; A38109.
CC TIGR; EF0799; -.
CC InterPro; IPR002901; Amidase_4.
CC InterPro; IPR002482; LysM.
CC Pfam; PF01832; Amidase_4; 1.
CC Pfam; PF01476; LysM; 5.
CC SMART; SM00257; LysM; 6.
CC SMART; SM00047; LY22; 1.
CC Cell division; Glycosidase; Bacteriolytic enzyme; Cell wall;
CC Hydrolase; Septation; Repeat; Signal; Complete proteome.
FT SIGNAL 1 53
FT CHAIN 54 737
FT REPEAT 363 405
FT REPEAT 431 473
FT REPEAT 499 541
FT REPEAT 567 609
FT REPEAT 633 675
FT REPEAT 695 737
FT REPEAT 85 85
FT CONFLICT 118 118
T -> I (IN REF. 1).
A -> V (IN REF. 1).

```

```

FT CONFLICT 143 143
FT CONFLICT 417 417
FT CONFLICT 449 449
FT CONFLICT 476 476
FT CONFLICT 484 484
FT CONFLICT 567 632
SQ SEQUENCE 737 AA; 77025 MW; ABB163D506AC7507 CRC64;

Query Match
Best Local Similarity 5.3%; Score 100; DB 1; Length 737;
Matches 59; Conservative 39; Mismatches 115; Indels 78; Gaps 15;

QY 41 RNESPKNKI-LNNHITAYSESHLYDRMTF-CLASHNTLNGACP-SENPSSSSV----- 94
DB 309 RYATDPSYNALNNVITAT--NLTDIDPSSCGNTGGTWNFGTGGSNQSGNTIYYTVK 366
QY 95 SGET--NITLQF-----TEKSL--IKRELQIKGYKQLAFKSYNCPISGLTLNSAHFNCNK 146
DB 367 SGTGLNKIAAQYGVSVANLRSWNGISGDLIFVGQKLIIVKKA---SGTGGSGSGSN 423
QY 147 AASGASLYIYPAGELKNIIPFGGIWDATLKLVRKARYSTGYTITINIKLTDKGNIQI 206
DB 424 -QSGTNTVYTVASGDTLN-----KIAAQYG-----VSVANLRS 455
QY 207 WLPOFKSDARVDLNLRLPTGGTYIGNSVDMCFYDGYSTNSSSLEIRFQDNRPKSDKPY 266
DB 456 W-----NGISGDL-----IFVGQK---LVKGAAGNTGGSNNGSNQSGNTIYY 499
QY 267 LRKINDTKEIA--YPLS-----LLLAGSLTPPINGTSLNIADAASLETNN 311
DB 500 TIKSGDTLNKIAAQYGVSVANLRSWNGISGDLIFAQKLIIVKGTSGNTGGSGSGSN 559
QY 312 R 312
DB 560 Q 560

RESULT 8
SYFB LISIN STANDARD; PRT; 802 AA.
ID -SYFB LISIN
AC Q92C16;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
DE (Phenylalanine-tRNA ligase beta chain) (PheRS).
GN PHE1 OR LIN185.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-Gel Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek E., Novella S., de Fabios B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + 5-phenylalanyl-tRNA(Phe).
CC -!- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

```

CC	-!	SIMILARITY:	Belongs to the phenylalanyl-tRNA synthetase beta chain family. Subfamily 1.
CC	-!	SIMILARITY:	Contains 1 tRNA-binding domain.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL;	AL596167;	CAC96416.1; -.
DR	PIR;	AH1580;	RH1580.
DR	ListList;	LIN01648;	-.
DR	HAWAP;	NF_00283;	--1
DR	InterPro;	IPR005146;	B3_4.
DR	InterPro;	IPR005147;	B5.
DR	InterPro;	IPR005121;	Fdx_AnticB.
DR	InterPro;	IPR008994;	Nucleic_acid_OB.
DR	InterPro;	IPR004532;	PheT_bact.
DR	InterPro;	IPR002547;	tRNA_bind.
DR	Pfam;	PF03483;	B3_4; 1.
DR	Pfam;	PF03484;	B5; 1.
DR	Pfam;	PF03147;	FDX-ACS; 1.
DR	Pfam;	PF01588;	tRNA_bind; 1.
DR	TIGRFAMS;	TIGR00472;	pheT_bact; 1.
DR	PROSITE;	PS50885;	TRBD; 1.
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Metal-binding; Magnesium; RNA-binding; tRNA-binding;		
KW	Complete proteome.		
KW	DOMAIN	40	155
FT	METAL	462	462
FT	METAL	468	468
FT	METAL	471	471
FT	METAL	472	472
FT	METAL	471	471
FT	METAL	472	472
SEQUENCE	802 AA;	88191 MW;	D75802193964D7D0 CRC64;
Query Match	5.2%;	Score 99;	DB 1; Length 802;
Best Local Similarity	17.3%;	Pred. No. 6.7;	Indels 140; Gaps 15;
Matches	70;	Conservative 63;	Mismatches 132;
QY	48	HNILNHITAYSESHLYDRMTFLCLSSHTNINCACPTSENPPSSSVSGETNLTQFTEK	107
DB	422	NRLGTEI-SLSSEITIFDLRGFLVEKEDTLIEVPTR-----RWDIIIE----	466
QY	108	RSLIKREIQIKGYKOLLFPKSNVCPSSLTNISMFNCNKNAASCASLIYVIPAGELKNLPF	167
DB	467	ADILEVARIVGYDRI--PVTLPATST-----	492
QY	168	GGIWDA TLKLVRKYRYSETYG-----TYTKN-----TIKLTDKNIQLWPQQFKSDARV	217
DB	493	GGLSDSQKARRVMRAYLEOAGUNQALTYSLTSSKKDATRLALSDEKTVALSPMPSEHSHL	552
QY	218	DNLRPF---TGCGTVGRNSVDMCFYD----GYSTNSSSLEIRFQ-----	255
DB	553	RFSIVEQLIRSASYNIAKRMMDVALYEMGTVFYATEGDNLPIEQEHLA GLTGNWHTADW	612
QY	256	DNPKPS-----DGKFYLKRINDDTKEIAYTSLLLIAGK-----SL	290
DB	613	QXTPKPDFVLKGIVEGLWNKLGIAELHWKIQEKSELHPORTASIQLEGKEIGYLGAL	672
QY	291	TTNGCTSLMIADAASLETNNKITATVMTPEISV-PVLCWE-----	329
DB	673	HFAVEASYLDKETVFEINVKALLDATKCKVVTHPLRPYPETMRDLLLVKDQTDHATIS	732
QY	330	-----GRLQIDAKVENPEAGYMGN-----INVTFTPSSOTL	361
DB	733	QVIKEHGGLLVDFIEDFEGESLGENKKSLEYTLTFLDSERTL	777
RESULT	9		

862 KE-NSHWELTGN---NVNQLMLTNG---HHHLNAQNDAN-----KV 896
 186 ETYGTYTINIITKTDKNIQIWLQPKSDA-RVDLNLRLPTGGTY-----IGR-NSVDM 237
 897 TYNTLTWN---SLSGNSFYVWDFTNKSKVNVKSAIGNFLQVADKTGEHNEL 953
 238 CFYDGYSTNSSLEIRFODNN-PSKDGKFLYRKINDTKEIATYLSLLAGKSLPTTNGT 296
 954 TLFDSATERNLEVLANGSVDRGAWKYKLRVNG--RYDLNYPEVEKRNQTVJTNIT 1011
 297 SUN--LADAASLETNWRITATMP 319
 1012 TPNDIQADAPSAQSNNEIARVETP 1036
 PRT; 608 AA.
 STANDARD;
 ID YD93 METJA
 AC Q58788;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1393.
 GN MJ1393.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervatage A.R., Dougherty B.A., Tomb J.P., Adams M.D., Reich C.L.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Gloeck A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.D., Nguyen D.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073 (1996).
 CC -1- SIMILARITY: TO MJANNASCHII MJ1394 AND A.FULGIDUS AF2028.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U67579; AAB99403.1; -;
 CC PIR; H64473; H64473.
 CC TIGR; MJ1393; -;
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSSEQ 4 24 POTENTIAL.
 SQ SEQUENCE 608 AA; 66768 MW; 010FAP1C29F9C73C CRC64;
 Query Match
 Best Local Similarity 20.3%; Pred.No.5.1;
 Matches 85; Conservative 68; Mismatches 119; Indels 147; Gaps 25;
 Qy 1 MNKILF---IPTLPSSVLFTFAVSADKIP-----GDESTNIFGRDR 41
 Db 1 MKKLIFALLMSLIFGTFVGY---GDNGPLVAYVYKYNITGNTGGLVSTI----- 52
 Qy 42 NESSPKNIINNHITAYSESHLYDRMTFLGLSSHNT-----LKG-----ACPTSE 87
 Db 53 -ESTGTIVINN--TGVTINDTLVDVVAVVAINSNITQPEVYVNGTPKGVISSAPAYT 109
 Qy 88 N-PSSSS-----VSGETNITQTEKRSIKRELQ-KGYKQLLFKSVNCPSGLTN-SA 139

110 KLPVNTYIHIPIIPNNSYVIKFAIDKSIITGVPLINE-----TYSDTKIFSERLSWSV 165
 140 HFNCKNNAASCASLYLYIPAGE-----LKNLPFGGIWDATLKURVKRISYETGT 189
 166 YLNISSNVSA-----LPATDTPVSVIMTKYLSNDP-----NNYGS 200
 190 YT---ENITIKLTGKNTQIW-----LPQFK-----SDARVDLNLRLPTGGT 228
 201 DTWNPNTTGAIANEGSITLWDGPFYLFEGYDLSLTWTGVTWNTKNVITINI--TGNT 258
 229 YIGRNSVDWCFYDGYSTNSSLEIRFODNNPKSDGKFLYRKINDTKEIATP-----LSL 283
 259 YNRTGTLMKY--GPAV-----TPFENGYSCTKI-----EGIVATGYGVS 300
 284 LIAGKSLPTNG-----TSLNIADAASLETNWRITATMPETI--SVPVLCWPORLQL 334
 301 TKEGPFNLASSGKYEIWTESANVSKAS--SYFNLTHVTIWAANGSFVILDPFNITL 357
 PRT; 758 AA.
 STANDARD;
 ID SP21 YEAST
 AC P35209;
 DT 01-FEB-1994 (Rel. 26, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE SPT21 protein.
 GN SPT21 OR YMR179W OR YN8010.09.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94186069; PubMed=8138180;
 RX Natsoulis G., Winston F., Boeke J.D.;
 RA "The SP110 and SPT21 genes of Saccharomyces cerevisiae.";
 RL Genetics 136:93-105 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313266; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagers K., Lye G., Moule S., Odeil C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 XII.";
 RL Nature 387:90-93 (1997).
 CC -1- FUNCTION: Required for normal transcription at a number of loci in
 CC yeast.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L24436; AAA5078.1; -;
 CC PIR; Z49808; CAA89912.1; -;
 CC GERMOnline; 142853; -;
 CC TRANSFAC; T04376; -;
 CC SGD; S0004751; SPT21.
 CC GO; GO:0006357; P:regulation of transcription from pol II pro. . .; IMP.
 CC DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).
 CC DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).
 CC SEQUENCE 758 AA; 84697 MW; 7DB3FCF7EE996705 CRC64;
 Query Match
 5.2%; Score 98; DB 1; Length 758;

DR GO:GO:0006406; P:mRNA-nucleus export; IMP.
DR GO:GO:0006939; P:nuclear pore organization and biogenesis; IMP.
DR InterPro: IPR007187; Nup133.
DR Pfam: PF04044; Nup133; 1
KW Nuclear protein; transport.
SQ SEQUENCE 1157 AA; 133319 MW; C8BDBB7D709C5C08 CRC64;

Query Match 5.2%; Score 98; DB 1; Length 1157;
Best Local Similarity 22.1%; Pred. No. 13;
Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

QY 43 ENILNNHITAY-----SESHLYDRMTF-----LCLSHNTLNACPTSENP 89
DB 96 YGLVNDHKVWNIHSTQKDTPTVTPFRSDNDDEIAVAPRCILTFATMDESEIALNP 155
QY 90 SSSSVSGE-----TNITQETKRSIKRELOIK-----GYKQLLFK 126
DB 155 NDQDETGGIIIGKKAIVYEDININNFKLSKFS---HELELPINSSGEXCDLM- 211
QY 127 SVNC-PSGLTNS-----AHFNCNKNAASGASLYLIPAGELXKLPG-GIWDATLKL- 178
DB 212 -LNCBPAGVLSTNMGRIFITIRNSMGKPKQLL---GKLNKPKFKLGIWSKIFNTNS 266
QY 179 -VKRYSETY--GTYTINITIKLTKGNIQIWLPOKSDARVDLNLRFPGGTYIGRNSV 235
DB 267 VVSLRNGPILGKGTRELVTIT--TNKGIFQTV--QLSA-----TNSHPT-----KLI 308
QY 235 DMCFYDGYSTNSSSL-----EIRQDNNPKSD--GKFLRKINDDT-KEIAYTLLSLLA 286
DB 309 DNVNIEAILESQDLYPFAHGLKWDHPLQDESSQLFSLSYDSSCNETYILLSTIF 368
QY 287 GKSLPTNGTSLNIADAASLETNNRITAVTM-PEISVP 324
DB 369 DSS-----SNSITSTVLTNFMESITDKPKPKTIF 402

RESULT 13
ID2 PAEMA
ID -CDG2 PAEMA STANDARD; PRT; 713 AA.
AC P31835;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 13-OCT-2003 (Rel. 42, Last annotation update)
DE Cyclomaltodextrin glucanotransferase precursor (EC 2.4.1.19)
DE (Cyclodextrin-glycosyltransferase) (CGTase).
OS Paenibacillus macerans (Bacillus macerans).
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX NCBI_TaxID=44252;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
RA Sugimoto T., Kubota M., Sakai S.;
RT polypeptide possessing cyclomaltodextrin glucanotransferase
RT activity.";
RL Patent number GB2169902, 23-JUL-1986.
CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
CC of a 1,4-alpha-D-glucosidic bond.
CC -1- COFACTOR: Binds 2 calcium ions per subunit (By similarity).
CC -1- SUBUNIT: Monomer.
CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE
CC IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND
CC IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER
CC ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN
CC ALPHA-1,4-GUCOSIDIC LINKAGE FOR CYCLIZING THE
CC MALTOOLIGOSACCHARIDE PRODUCED.
CC -1- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
DR PIR; S26589; ALESSR.
DR HSSP; P43379; ICDG.
DR InterPro: IPR006589; Alp_amy1_cat_sub.
DR InterPro: IPR006048; Alpha_amy1_C-
DR InterPro: IPR006047; Alpha_amy1_cat.
DR InterPro: IPR002044; CBD_4.
DR InterPro: IPR006046; glyco_hydro__3.
DR InterPro: IPR007110; Ig-like.

```
DR InterPro; IPR002909; IPT TIG.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02806; alpha-amylase_C; 1.
DR Pfam; PF02886; CBM_20; 1.
DR Pfam; PF01833; TIG; 1.
DR PRINTS; PRC0110; ALPHAAMYLASE.
DR ProDom; PD001568; CBD_4; 1.
DR SMART; SM00632; AmyC; 1.
DR TRANSFase; Glycosyltransferase; Calcium-binding; Signal.
FT SIGNAL 1 27 CYCLOMALTODEXTRIN GLUCANOTRANSFERASE.
FT CHAIN 28 713 A1.
FT DOMAIN 28 165 B.
FT DOMAIN 166 229 B.
FT DOMAIN 230 434 A2.
FT DOMAIN 435 522 C.
FT DOMAIN 523 609 D.
FT DOMAIN 610 713 E.
FT ACT_SITE 256 256 BY SIMILARITY.
FT ACT_SITE 285 285 BY SIMILARITY.
FT ACT_SITE 356 356 BY SIMILARITY.
FT METAL 54 54 CALCIUM 2 (BY SIMILARITY).
FT METAL 56 56 CALCIUM 2 (VIA CARBONYL OXYGEN) (BY SIMILARITY).
FT METAL 59 59 CALCIUM 2 (BY SIMILARITY).
FT METAL 60 60 CALCIUM 2 (BY SIMILARITY).
FT METAL 78 78 CALCIUM 2 (VIA CARBONYL OXYGEN) (BY SIMILARITY).
FT METAL 80 80 CALCIUM 2 (BY SIMILARITY).
FT METAL 156 156 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY SIMILARITY).
FT METAL 217 217 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY SIMILARITY).
FT METAL 226 226 CALCIUM 1 (BY SIMILARITY).
FT METAL 260 260 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY SIMILARITY).
SQ SEQUENCE 713 AA; 76857 MW; 5A287BCC4A4PE35 CRC64;

Query Match 5.1%; Score 95.5; DB 1; Length 713;
Best Local Similarity 20.1%; Pred. No. 11;
Matches 65; Conservative 43; Mismatches 131; Indels 85; Gaps 12;

QY 68 MTFICLSSHTNLGACPTSENPPSSSVSGETNITLQTEKRSLIKRELQKGYKQLLFKS 127
D 8 LTVSVMSVGIALGAALPWAAP-DTSVNNKLNFSG---TDTVYQIVTDFVDG-----NS 57
QY 128 VNCPSGHTLSAHNCKNKAASGLSLYLIPAGELXNLPFGGIWDATLKLVRKRYSEY 187
D 58 ANNPFGAFAFSDHGN-----LKY-----FGDNQG-----ITNKINDGY 92
QY 186 GTYVINITIKLTKENQIWLPPQFKSDARVDLNLRLPTGGTYIG----- 231
D 93 -----LTGNGITALWISQPVENITAVINYSVGNVNTAYHGYWPRDFKKTNAAFGSF 142
QY 232 -----RNSVDMCFYDGYSTNSLSLEIRFQDNPKSDGKFVLRKINDDTKEIA 278
D 143 TDFGNLAAASHNLIKVVWDFAPNHTNPASSTDFSAENGALYNNGTLLGYSNDTAGLF 202
QY 279 YTLISLLAGKSLTPT-NGTSNIADAASLETNWRNRTAVTWPETISVPVLCWPGRLQDAK 337
D 203 HH-----NGGTFPTTESGIYKNLYDLADINQNNNTIDSYLKESQLWNLGVGIRFDV 258
QY 338 VENPRAGQ--YMGNT-----NVITPT 355
D 259 KEMPOGWQKSVSSSIYSSANPVET 282

RESULT 14
REFM YEAST
ID 3F1M YEAST STANDARD; PRT; 413 AA.
AC P30775;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

Peptide chain release factor 1, mitochondrial precursor (MRF-1).
MRF1 OR YGLI43C.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A. PubMed=1475194;
MEDLINE=93111710;
Pel H.J., Maat M.J., Rep M., Grivell L.A.;
"The yeast nuclear gene MRF1 encodes a mitochondrial peptide chain
release factor and cures several mitochondrial RNA splicing
defects.";
Nucleic Acids Res. 20:6339-6346(1992).
[2]
SEQUENCE FROM N.A.
STRAIN=S288C / FY1769;
MEDLINE=97197983; PubMed=9046099;
Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
"The sequence of a nearly unclonable 22.8 kb segment on the left arm
of chromosome VII from Saccharomyces cerevisiae reveals ARO2, RLSA,
TRP1, MRF1 genes and six new open reading frames.";
Yeast 13:177-182(1997).
-!- FUNCTION: Mitochondrial peptide chain release factor that directs
the termination of translation in response to the peptide chain
termination codons UAA and UAG.
-!- SUBCELLULAR LOCATION: Mitochondrial.
-!- SIMILARITY: Belongs to the prokaryotic/mitochondrial release
factor family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; X60381; CAA42932.1; -
EMBL; X99960; CAA68219.1; -
EMBL; Z72665; CAA96855.1; -
PIR; S28602; S28602.
GeneOnline; 141191.
SGD; S000311; MRF1.
GO; GO:0005739; C:mitochondrion; IDA. factor activity; IMP.
GO; GO:0003747; F:translation release factor activity; IMP.
GO; GO:0006415; P:translational termination; IMP.
InterPro; IPR005139; PCRF.
InterPro; IPR000352; Pep_rel_factor_1.
Pfam; PF03452; PCRF; 1.
Pfam; PF00472; RF-1; 1.
PROSITE; PS00745; RF_PROK_1; 1.
Protein biosynthesis; Mitochondrion; Transit peptide.
TRANSIT ? ? MITOCHONDRION (POTENTIAL).
CHAIN ? 413 PEPTIDE CHAIN RELEASE FACTOR 1.
SEQUENCE 413 AA; 46770 MW; AAAD4829748C7604 CRC64;

Query Match 5.0%; Score 95; DB 1; Length 413;
Best Local Similarity 21.8%; Pred. No. 5.7;
Matches 67; Conservative 47; Mismatches 121; Indels 72; Gaps 14;

QY 30 SSSSVSGETN--ITLQTEKRSLI-----KRELQKGYKQLLFKSVNCPGSLTNSAHFN 142
D 30 TSTTNSKNGSIPTQYTELSPLLVKAQKVEALKD-----LDRKLSG-----GIHFD 78
QY 143 CNK-----NAASGASLYLYIPAGELXNLPFGGIWDATLKLVRKRYSE---TYGIYTI 192
D 79 VNKQKHYAKLSALDTTIEYKELNELKSLQEMIVSDPSLRAAEQEYAEILVPPQYETSS 238
QY 193 NITIKLTKGNIQIWLPOFKSDARVDLNLRLPTGG---TYIGRNSVDMCFYDGYSTNSSS 249
D 139 RLVNKLPLP-----PEPFADKPSLLELRGVGGVIEAMITQNLDM--YIGYA-NYRK 187
```

QV 250 LEIRFQDNPKSDGKFLRKINDTKEIAYTSLLLACK-----SITPTNGTSINTADA 303
Dd 188 WKYILISKVENESGGIIDALLSIEEAGSYDLRPFAGVHRVQIPSTETRG----- 239
QV 304 ASLTNNNRITAVTMPEI-----SVPLVCPGRQLQDAKVENPEAGVMGN---1 350
Dd 240 ---RTHSTAATAVVLPLQIGDESAKSIDAYETFRKGEIRVDIMRASGGGGHVTTSADV 296
QV 351 NVITPES 357
Dd 297 RLTHIPS 303
RESULT 15
WAPA_BACSU
ID WAPA_BACSU STANDARD; PRT; 2334 AA.
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G OR BSU39230.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT Protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Miwa Y., Fujita Y.;
RT Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lic and cel loci, and creation of a 177 kb contig
RT covering the gnt-sacxy region."
RL Microbiology 142:3113-3123(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96044033; PubMed=9384377;
RA Kunst F., Ogasawara K., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brigrell S.C., Bron S.,
RA Brucillet S., Bruschi C.V., Caldwell B., Capuano N.J., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.J., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech O., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Ketterer P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lepidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda C., Mael C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogiwara K., Oudega A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purrelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger N., Rivoita C., Rocha E., Roche S., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemart K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toseato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger C.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E.COLI Rhs GROUP OF PROTEINS (RhsA-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by not-for-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29885; BAA06260.1; -
DR EMBL; D83026; BAA11693.1; -
DR EMBL; Z99124; CAB15959.1; -
DR FIC; S32920; S32920.
DR Subtilisin; BG10797; wapa.
DR InterPro; IPR003305; CBM_CenC.
DR InterPro; IPR006530; YD.
DR Pfam; PF02018; CBM_4_9; 1.
DR Pfam; PF05893; Rhs_repeat; 14.
DR TIGRfam; TIGR01643; YD_repeat; 2x; 17.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 2334
FT DOMAIN 504 869
FT REPEAT 504 605
FT REPEAT 636 736
FT REPEAT 769 869
FT DOMAIN 1021 2139
FT REPEAT 1021 1040
FT REPEAT 1042 1061
FT REPEAT 1063 1082
FT REPEAT 1083 1102
FT REPEAT 1109 1128
FT REPEAT 1129 1148
FT REPEAT 1150 1169
FT REPEAT 1174 1193
FT REPEAT 1199 1218
FT REPEAT 1219 1238
FT REPEAT 1238 1258
FT REPEAT 1258 1278
FT REPEAT 1278 1298
FT REPEAT 1298 1318
FT REPEAT 1318 1338
FT REPEAT 1338 1358
FT REPEAT 1358 1378
FT REPEAT 1378 1398
FT REPEAT 1398 1418
FT REPEAT 1418 1438
FT REPEAT 1438 1458
FT REPEAT 1458 1478
FT REPEAT 1478 1498
FT REPEAT 1498 1518
FT REPEAT 1518 1538
FT REPEAT 1538 1558
FT REPEAT 1558 1578
FT REPEAT 1578 1598
FT REPEAT 1598 1618
FT REPEAT 1618 1638
FT REPEAT 1638 1658
FT REPEAT 1658 1678
FT REPEAT 1678 1698
FT REPEAT 1698 1718
FT REPEAT 1718 1738
FT REPEAT 1738 1758
FT REPEAT 1758 1778
FT REPEAT 1778 1798
FT REPEAT 1798 1818
FT REPEAT 1818 1839
FT REPEAT 1839 1859
FT REPEAT 1859 1879
FT REPEAT 1879 1899
FT REPEAT 1899 1919
FT REPEAT 1919 1939
FT REPEAT 1939 1959
FT REPEAT 1959 1979
FT REPEAT 1979 1999
FT REPEAT 1999 2019
FT REPEAT 2019 2039
FT REPEAT 2039 2059
FT REPEAT 2059 2079
FT REPEAT 2079 2099
FT REPEAT 2099 2119
FT REPEAT 2119 2139
FT REPEAT 2139 2159
FT REPEAT 2159 2179
FT REPEAT 2179 2199
FT REPEAT 2199 2219
FT REPEAT 2219 2239
FT REPEAT 2239 2259
FT REPEAT 2259 2279
FT REPEAT 2279 2299
FT REPEAT 2299 2319
FT REPEAT 2319 2334
FT REPEAT 2334 2354
FT REPEAT 2354 2374
FT REPEAT 2374 2394
FT REPEAT 2394 2414
FT REPEAT 2414 2434
FT REPEAT 2434 2454
FT REPEAT 2454 2474
FT REPEAT 2474 2494
FT REPEAT 2494 2514
FT REPEAT 2514 2534
FT REPEAT 2534 2554
FT REPEAT 2554 2574
FT REPEAT 2574 2594
FT REPEAT 2594 2614
FT REPEAT 2614 2634
FT REPEAT 2634 2654
FT REPEAT 2654 2674
FT REPEAT 2674 2694
FT REPEAT 2694 2714
FT REPEAT 2714 2734
FT REPEAT 2734 2754
FT REPEAT 2754 2774
FT REPEAT 2774 2794
FT REPEAT 2794 2814
FT REPEAT 2814 2834
FT REPEAT 2834 2854
FT REPEAT 2854 2874
FT REPEAT 2874 2894
FT REPEAT 2894 2914
FT REPEAT 2914 2934
FT REPEAT 2934 2954
FT REPEAT 2954 2974
FT REPEAT 2974 2994
FT REPEAT 2994 3014
FT REPEAT 3014 3034
FT REPEAT 3034 3054
FT REPEAT 3054 3074
FT REPEAT 3074 3094
FT REPEAT 3094 3113
FT REPEAT 3113 3133
FT REPEAT 3133 3153
FT REPEAT 3153 3173
FT REPEAT 3173 3193
FT REPEAT 3193 3213
FT REPEAT 3213 3233
FT REPEAT 3233 3253
FT REPEAT 3253 3273
FT REPEAT 3273 3293
FT REPEAT 3293 3313
FT REPEAT 3313 3333
FT REPEAT 3333 3353
FT REPEAT 3353 3373
FT REPEAT 3373 3393
FT REPEAT 3393 3413
FT REPEAT 3413 3433
FT REPEAT 3433 3453
FT REPEAT 3453 3473
FT REPEAT 3473 3493
FT REPEAT 3493 3513
FT REPEAT 3513 3533
FT REPEAT 3533 3553
FT REPEAT 3553 3573
FT REPEAT 3573 3593
FT REPEAT 3593 3613
FT REPEAT 3613 3633
FT REPEAT 3633 3653
FT REPEAT 3653 3673
FT REPEAT 3673 3693
FT REPEAT 3693 3713
FT REPEAT 3713 3733
FT REPEAT 3733 3753
FT REPEAT 3753 3773
FT REPEAT 3773 3793
FT REPEAT 3793 3813
FT REPEAT 3813 3833
FT REPEAT 3833 3853
FT REPEAT 3853 3873
FT REPEAT 3873 3893
FT REPEAT 3893 3913
FT REPEAT 3913 3933
FT REPEAT 3933 3953
FT REPEAT 3953 3973
FT REPEAT 3973 3993
FT REPEAT 3993 4013
FT REPEAT 4013 4033
FT REPEAT 4033 4053
FT REPEAT 4053 4073
FT REPEAT 4073 4093
FT REPEAT 4093 4113
FT REPEAT 4113 4133
FT REPEAT 4133 4153
FT REPEAT 4153 4173
FT REPEAT 4173 4193
FT REPEAT 4193 4213
FT REPEAT 4213 4233
FT REPEAT 4233 4253
FT REPEAT 4253 4273
FT REPEAT 4273 4293
FT REPEAT 4293 4313
FT REPEAT 4313 4333
FT REPEAT 4333 4353
FT REPEAT 4353 4373
FT REPEAT 4373 4393
FT REPEAT 4393 4413
FT REPEAT 4413 4433
FT REPEAT 4433 4453
FT REPEAT 4453 4473
FT REPEAT 4473 4493
FT REPEAT 4493 4513
FT REPEAT 4513 4533
FT REPEAT 4533 4553
FT REPEAT 4553 4573
FT REPEAT 4573 4593
FT REPEAT 4593 4613
FT REPEAT 4613 4633
FT REPEAT 4633 4653
FT REPEAT 4653 4673
FT REPEAT 4673 4693
FT REPEAT 4693 4713
FT REPEAT 4713 4733
FT REPEAT 4733 4753
FT REPEAT 4753 4773
FT REPEAT 4773 4793
FT REPEAT 4793 4813
FT REPEAT 4813 4833
FT REPEAT 4833 4853
FT REPEAT 4853 4873
FT REPEAT 4873 4893
FT REPEAT 4893 4913
FT REPEAT 4913 4933
FT REPEAT 4933 4953
FT REPEAT 4953 4973
FT REPEAT 4973 4993
FT REPEAT 4993 5013
FT REPEAT 5013 5033
FT REPEAT 5033 5053
FT REPEAT 5053 5073
FT REPEAT 5073 5093
FT REPEAT 5093 5113
FT REPEAT 5113 5133
FT REPEAT 5133 5153
FT REPEAT 5153 5173
FT REPEAT 5173 5193
FT REPEAT 5193 5213
FT REPEAT 5213 5233
FT REPEAT 5233 5253
FT REPEAT 5253 5273
FT REPEAT 5273 5293
FT REPEAT 5293 5313
FT REPEAT 5313 5333
FT REPEAT 5333 5353
FT REPEAT 5353 5373
FT REPEAT 5373 5393
FT REPEAT 5393 5413
FT REPEAT 5413 5433
FT REPEAT 5433 5453
FT REPEAT 5453 5473
FT REPEAT 5473 5493
FT REPEAT 5493 5513
FT REPEAT 5513 5533
FT REPEAT 5533 5553
FT REPEAT 5553 5573
FT REPEAT 5573 5593
FT REPEAT 5593 5613
FT REPEAT 5613 5633
FT REPEAT 5633 5653
FT REPEAT 5653 5673
FT REPEAT 5673 5693
FT REPEAT 5693 5713
FT REPEAT 5713 5733
FT REPEAT 5733 5753
FT REPEAT 5753 5773
FT REPEAT 5773 5793
FT REPEAT 5793 5813
FT REPEAT 5813 5833
FT REPEAT 5833 5853
FT REPEAT 5853 5873
FT REPEAT 5873 5893
FT REPEAT 5893 5913
FT REPEAT 5913 5933
FT REPEAT 5933 5953
FT REPEAT 5953 5973
FT REPEAT 5973 5993
FT REPEAT 5993 6013
FT REPEAT 6013 6033
FT REPEAT 6033 6053
FT REPEAT 6053 6073
FT REPEAT 6073 6093
FT REPEAT 6093 6113
FT REPEAT 6113 6133
FT REPEAT 6133 6153
FT REPEAT 6153 6173
FT REPEAT 6173 6193
FT REPEAT 6193 6213
FT REPEAT 6213 6233
FT REPEAT 6233 6253
FT REPEAT 6253 6273
FT REPEAT 6273 6293
FT REPEAT 6293 6313
FT REPEAT 6313 6333
FT REPEAT 6333 6353
FT REPEAT 6353 6373
FT REPEAT 6373 6393
FT REPEAT 6393 6413
FT REPEAT 6413 6433
FT REPEAT 6433 6453
FT REPEAT 6453 6473
FT REPEAT 6473 6493
FT REPEAT 6493 6513
FT REPEAT 6513 6533
FT REPEAT 6533 6553
FT REPEAT 6553 6573
FT REPEAT 6573 6593
FT REPEAT 6593 6613
FT REPEAT 6613 6633
FT REPEAT 6633 6653
FT REPEAT 6653 6673
FT REPEAT 6673 6693
FT REPEAT 6693 6713
FT REPEAT 6713 6733
FT REPEAT 6733 6753
FT REPEAT 6753 6773
FT REPEAT 6773 6793
FT REPEAT 6793 6813
FT REPEAT 6813 6833
FT REPEAT 6833 6853
FT REPEAT 6853 6873
FT REPEAT 6873 6893
FT REPEAT 6893 6913
FT REPEAT 6913 6933
FT REPEAT 6933 6953
FT REPEAT 6953 6973
FT REPEAT 6973 6993
FT REPEAT 6993 7013
FT REPEAT 7013 7033
FT REPEAT 7033 7053
FT REPEAT 7053 7073
FT REPEAT 7073 7093
FT REPEAT 7093 7113
FT REPEAT 7113 7133
FT REPEAT 7133 7153
FT REPEAT 7153 7173
FT REPEAT 7173 7193
FT REPEAT 7193 7213
FT REPEAT 7213 7233
FT REPEAT 7233 7253
FT REPEAT 7253 7273
FT REPEAT 7273 7293
FT REPEAT 7293 7313
FT REPEAT 7313 7333
FT REPEAT 7333 7353
FT REPEAT 7353 7373
FT REPEAT 7373 7393
FT REPEAT 7393 7413
FT REPEAT 7413 7433
FT REPEAT 7433 7453
FT REPEAT 7453 7473
FT REPEAT 7473 7493
FT REPEAT 7493 7513
FT REPEAT 7513 7533
FT REPEAT 7533 7553
FT REPEAT 7553 7573
FT REPEAT 7573 7593
FT REPEAT 7593 7613
FT REPEAT 7613 7633
FT REPEAT 7633 7653
FT REPEAT 7653 7673
FT REPEAT 7673 7693
FT REPEAT 7693 7713
FT REPEAT 7713 7733
FT REPEAT 7733 7753
FT REPEAT 7753 7773
FT REPEAT 7773 7793
FT REPEAT 7793 7813
FT REPEAT 7813 7833
FT REPEAT 7833 7853
FT REPEAT 7853 7873
FT REPEAT 7873 7893
FT REPEAT 7893 7913
FT REPEAT 7913 7933
FT REPEAT 7933 7953
FT REPEAT 7953 7973
FT REPEAT 7973 7993
FT REPEAT 7993 8013
FT REPEAT 8013 8033
FT REPEAT 8033 8053
FT REPEAT 8053 8073
FT REPEAT 8073 8093
FT REPEAT 8093 8113
FT REPEAT 8113 8133
FT REPEAT 8133 8153
FT REPEAT 8153 8173
FT REPEAT 8173 8193
FT REPEAT 8193 8213
FT REPEAT 8213 8233
FT REPEAT 8233 8253
FT REPEAT 8253 8273
FT REPEAT 8273 8293
FT REPEAT 8293 8313
FT REPEAT 8313 8333
FT REPEAT 8333 8353
FT REPEAT 8353 8373
FT REPEAT 8373 8393
FT REPEAT 8393 8413
FT REPEAT 8413 8433
FT REPEAT 8433 8453
FT REPEAT 8453 8473
FT REPEAT 8473 8493
FT REPEAT 8493 8513
FT REPEAT 8513 8533
FT REPEAT 8533 8553
FT REPEAT 8553 8573
FT REPEAT 8573 8593
FT REPEAT 8593 8613
FT REPEAT 8613 8633
FT REPEAT 8633 8653
FT REPEAT 8653 8673
FT REPEAT 8673 8693
FT REPEAT 8693 8713
FT REPEAT 8713 8733
FT REPEAT 8733 8753
FT REPEAT 8753 8773
FT REPEAT 8773 8793
FT REPEAT 8793 8813
FT REPEAT 8813 8833
FT REPEAT 8833 8853
FT REPEAT 8853 8873
FT REPEAT 8873 8893
FT REPEAT 8893 8913
FT REPEAT 8913 8933
FT REPEAT 8933 8953
FT REPEAT 8953 8973
FT REPEAT 8973 8993
FT REPEAT 8993 9013
FT REPEAT 9013 9033
FT REPEAT 9033 9053
FT REPEAT 9053 9073
FT REPEAT 9073 9093
FT REPEAT 9093 9113
FT REPEAT 9113 9133
FT REPEAT 9133 9153
FT REPEAT 9153 9173
FT REPEAT 9173 9193
FT REPEAT 9193 9213
FT REPEAT 9213 9233
FT REPEAT 9233 9253
FT REPEAT 9253 9273
FT REPEAT 9273 9293
FT REPEAT 9293 9313
FT REPEAT 9313 9333
FT REPEAT 9333 9353
FT REPEAT 9353 9373
FT REPEAT 9373 9393
FT REPEAT 9393 9413
FT REPEAT 9413 9433
FT REPEAT 9433 9453
FT REPEAT 9453 9473
FT REPEAT 9473 9493
FT REPEAT 9493 9513
FT REPEAT 9513 9533
FT REPEAT 9533 9553
FT REPEAT 9553 9573
FT REPEAT 9573 9593
FT REPEAT 9593 9613
FT REPEAT 9613 9633
FT REPEAT 9633 9653
FT REPEAT 9653 9673
FT REPEAT 9673 9693
FT REPEAT 9693 9713
FT REPEAT 9713 9733
FT REPEAT 9733 9753
FT REPEAT 9753 9773
FT REPEAT 9773 9793
FT REPEAT 9793 9813
FT REPEAT 9813 9833
FT REPEAT 9833 9853
FT REPEAT 9853 9873
FT REPEAT 9873 9893
FT REPEAT 9893 9913
FT REPEAT 9913 9933
FT REPEAT 9933 9953
FT REPEAT 9953 9973
FT REPEAT 9973 9993
FT REPEAT 9993 10013
FT REPEAT 10013 10033
FT REPEAT 10033 10053
FT REPEAT 10053 10073
FT REPEAT 10073 10093
FT REPEAT 10093 10113
FT REPEAT 10113 10133
FT REPEAT 10133 10153
FT REPEAT 10153 10173
FT REPEAT 10173 10193
FT REPEAT 10193 10213
FT REPEAT 10213 10233
FT REPEAT 10233 10253
FT REPEAT 10253 10273
FT REPEAT 10273 10293
FT REPEAT 10293 10313
FT REPEAT 10313 10333
FT REPEAT 10333 10353
FT REPEAT 10353 10373
FT REPEAT 10373 10393
FT REPEAT 10393 10413
FT REPEAT 10413 10433
FT REPEAT 10433 10453
FT REPEAT 10453 10473
FT REPEAT 10473 10493
FT REPEAT 10493 10513
FT REPEAT 10513 10533
FT REPEAT 10533 10553
FT REPEAT 10553 10573
FT REPEAT 10573 10593
FT REPEAT 10593 10613
FT REPEAT 10613 10633
FT REPEAT 10633 10653
FT REPEAT 10653 10673
FT REPEAT 10673 10693
FT REPEAT 10693 10713
FT REPEAT 10713 10733
FT REPEAT 10733 10753
FT REPEAT 10753 10773
FT REPEAT 10773 10793
FT REPEAT 10793 10813
FT REPEAT 10813 10833
FT REPEAT 10833 10853
FT REPEAT 10853 10873
FT REPEAT 10873 10893
FT REPEAT 10893 10913
FT REPEAT 10913 10933
FT REPEAT 10933 10953
FT REPEAT 10953 10973
FT REPEAT 10973 10993
FT REPEAT 10993 11013
FT REPEAT 11013 11033
FT REPEAT 11033 11053
FT REPEAT 11053 11073
FT REPEAT 11073 11093
FT REPEAT 11093 11113
FT REPEAT 11113 11133
FT REPEAT 11133 11153
FT REPEAT 11153 11173
FT REPEAT 11173 11193
FT REPEAT 11193 11213
FT REPEAT 11213 11233
FT REPEAT 11233 11253
FT REPEAT 11253 11273
FT REPEAT 11273 11293
FT REPEAT 11293 11313
FT REPEAT 11313 11333
FT REPEAT 11333 11353
FT REPEAT 11353 11373
FT REPEAT 11373 11393
FT REPEAT 11393 11413
FT REPEAT 11413 11433
FT REPEAT 11433 11453
FT REPEAT 11453 11473
FT REPEAT 11473 11493
FT REPEAT 11493 11513
FT REPEAT 11513 11533
FT REPEAT 11533 11553
FT REPEAT 11553 11573
FT REPEAT 11573 11593
FT REPEAT 11593 11613
FT REPEAT 11613 11633
FT REPEAT 11633 11653
FT REPEAT 11653 11673
FT REPEAT 11673 11693
FT REPEAT 11693 11713
FT REPEAT 11713 11733
FT REPEAT 11733 11753
FT REPEAT 11753 11773
FT REPEAT 11773 11793
FT REPEAT 11793 11813
FT REPEAT 11813 11833
FT REPEAT 11833 11853
FT REPEAT 11853 11873
FT REPEAT 11873 11893
FT REPEAT 11893 11913
FT REPEAT 11913 11933
FT REPEAT 11933 11953
FT REPEAT 11953 11973
FT REPEAT 11973 11993
FT REPEAT 11993 12013
FT REPEAT 12013 12033
FT REPEAT 12033 12053
FT REPEAT 12053 12073
FT REPEAT 12073 12093
FT REPEAT 12093 12113
FT REPEAT 12113 12133
FT REPEAT 12133 12153
FT REPEAT 12153 12173
FT REPEAT 12173 12193
FT REPEAT 12193 12213
FT REPEAT 12213 12233
FT REPEAT 12233 12253
FT REPEAT 12253 12273
FT REPEAT 12273 12293
FT REPEAT 12293 12313
FT REPEAT 12313 12333
FT REPEAT 12333 12353
FT REPEAT 12353 12373
FT REPEAT 12373 12393
FT REPEAT 12393 12413
FT REPEAT 12413 12433
FT REPEAT 12433 12453
FT REPEAT 12453 12473
FT REPEAT 12473 12493
FT REPEAT 12493 12513
FT REPEAT 12513 12533
FT REPEAT 12533 12553
FT REPEAT 12553 12573
FT REPEAT 12573 12593
FT REPEAT 12593 12613
FT REPEAT 12613 12633
FT REPEAT 12633 12653
FT REPEAT 12653 12673
FT REPEAT 12673 12693
FT REPEAT 12693 12713
FT REPEAT 12713 12733
FT REPEAT 12733 12753
FT REPEAT 12753 12773
FT REPEAT 12773 12793
FT REPEAT 12793 12813
FT REPEAT 12813 12833
FT REPEAT 12833 12853
FT REPEAT 12853 12873
FT REPEAT 12873 12893
FT REPEAT 12893 12913
FT REPEAT 12913 12933
FT REPEAT 12933 12953
FT REPEAT 12953 12973
FT REPEAT 12973 12993
FT REPEAT 12993 13013
FT REPEAT 13013 13033
FT REPEAT 13033 13053
FT REPEAT 13053 13073
FT REPEAT 13073 13093
FT REPEAT 13093 13113
FT REPEAT 13113 13133
FT REPEAT 13133 13153
FT REPEAT 13153 13173
FT REPEAT 13173 13193
FT REPEAT 13193 13213
FT REPEAT 13213 13233
FT REPEAT 13233 13253
FT REPEAT 13253 13273
FT REPEAT 13273 13293
FT REPEAT 13293 13313
FT REPEAT 13313 13333
FT REPEAT 13333 13353
FT REPEAT 13353 13373
FT REPEAT 13373 13393
FT REPEAT 13393 13413
FT REPEAT 13413 13433
FT REPEAT 13433 13453
FT REPEAT 13453 13473
FT REPEAT 13473 13493
FT REPEAT 13493 13513
FT REPEAT 13513 13533
FT REPEAT 13533 13553
FT REPEAT 13553 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: May 6, 2004, 10:16:37 ; Search time 23 Seconds
(without alignments)
810.303 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1888
Sequence: 1 MNKILFIFLTFSSVLTFPA.....EAGQYMGKINVTFTFSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCITUS COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17
2	988.5	52.4	363	2	US-08-483-101-16
3	927.5	49.2	364	2	US-08-483-101-5
4	116.5	6.2	1600	2	US-08-617-697-10
5	114	6.0	650	3	US-08-362-525-2
6	111.5	5.9	1035	4	US-09-206-942-41
7	111.5	5.9	1011	4	US-09-206-942-39
8	111.5	5.9	1529	2	US-08-728-470-10
9	111.5	5.9	1529	3	US-08-719-641-10
10	108	5.7	524	4	US-09-242-913B-15
11	107	5.7	1095	4	US-09-206-942-45
12	107	5.7	1101	4	US-09-206-942-43
13	106	5.6	915	4	US-08-737-716-13
14	101	5.4	971	4	US-09-206-942-35
15	101	5.4	1222	4	US-09-206-942-37
16	101	5.4	1228	4	US-09-206-942-34
17	101	5.4	1338	2	US-08-728-470-9
18	101	5.4	1338	3	US-08-719-641-9
19	101	5.4	1599	2	US-08-617-697-3
20	100.5	5.3	1004	4	US-09-206-942-57
21	100.5	5.3	1010	4	US-09-206-942-55
22	99.5	5.3	969	4	US-09-206-942-32
23	99.5	5.3	975	4	US-09-206-942-30
24	99	5.2	992	4	US-09-206-942-61
25	99	5.2	998	4	US-09-206-942-59
26	99	5.2	1848	3	US-08-236-791-6
27	99	5.2	1848	4	US-08-839-996-6

28	59	5.2	1848	4	US-10-080-505-6	Sequence 6, Appli
29	59	5.2	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
30	97	5.1	977	4	US-09-206-942-53	Sequence 53, Appl
31	97	5.1	983	4	US-09-206-942-51	Sequence 51, Appl
32	94.5	5.0	843	4	US-09-543-681A-6880	Sequence 6880, Ap
33	94.5	5.0	1912	1	US-08-409-995-4	Sequence 4, Appli
34	94.5	5.0	1912	3	US-08-685-457-4	Sequence 4, Appli
35	94.5	5.0	2353	3	US-09-377-155-33	Sequence 33, Appl
36	94.5	5.0	2353	3	US-08-913-942-4	Sequence 4, Appli
37	94.5	5.0	2353	4	US-09-669-974-33	Sequence 33, Appl
38	94.5	5.0	2353	4	US-09-797-862-33	Sequence 47, Appl
39	94.5	5.0	2354	4	US-09-268-347-47	Sequence 36, Appl
40	94.5	5.0	2411	4	US-09-268-347-36	Sequence 134, App
41	93.5	5.0	715	4	US-09-462-917A-134	Sequence 5351, Ap
42	93.5	5.0	901	4	US-09-134-001C-5351	Sequence 6, Appli
43	93	4.9	1167	1	US-08-485-568A-6	Sequence 6, Appli
44	93	4.9	1167	2	US-08-590-554A-6	Sequence 6, Appli
45	93	4.9	1167	2	US-09-184-223-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-483-101-17
; Sequence 17, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Feiber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-483-101-17

Query Match 91.5%; Score 1536.5; DB 2; Length 360;
Best Local Similarity 50.2%; Pred. No. 1.3e-149;
Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;
QY 1 MNKILFIFLTFSSVLTFPAVSADKIPGDSINIFGPRDRNESSPKHNLNHTAYSE 60
Db 1 MNKILFIFLTFSSVLTFPAVSADKIPGDSINIFGPRDRNESSPKHNLNHTAYNG 60

[illegible]

```

RESULT 2
US-08-483-101-16
; Sequence 16, Application US/05483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Cazor, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8090
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
US-08-483-101-16

```

```

QY 3 KILFIETLFPSSVLFPFAVSADKIPGDE--SITNIF-GPR-DRNESSPKHNLNNHITAY 58
DB 1 :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 57
2 KKIETL-----SIIPSAVAGRYPETVGNLTKSFQAPRLDRSVQSPFIYIFTNHVAGY
QY 59 SEBHTLYDRMTFLCLSSHNTLNGACPTSENPSSSVS-GETN-TLOFTEKESLIKRELIQI 117
DB 1 :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 114
58 SLBHSLYDRIVFLCTSSNPVNGACPFI--GTSGVQVGTITITLOFTEKESLIKENIL 174
DB 1 :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 174
115 AGNKKPIWENQODFSNLVLNLSKSWCSGAGNANGTLLNIYIPAGEINKLPTFGGIWEAT 174
QY 175 LKLVKERYSET---YGTVTINITIKLTDKGNITQILWLPQKSDARVDLNLRPCTGGTYI 230
DB 1 :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 233
175 LILKLS-RYGEVSTHYGNYIVNTIVDLTDKGNIQVWLPGGHSPRVLDNLRPIONYKYS 233
QY 231 GRNSVDMCFYDGYSTNSSSLEIRPDNNPKDGKFLFKINDDTKEIAYTJSLLLAGKSL 290
DB 1 :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 292
234 GSNLWCMCFYDGYSTNDSVVIKQDCNPTNSSEYNLYKIG-GTEKLPYAVSLLMGEKIF
QY 251 TPTNGTS-NTADAASLETNNRITAVTMPEISVFLCWPGRLDLDKAVENPEAGQYMGNI 350
DB 1 :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 352
253 YPVNQCQSTINDSSVLETNMNRVTAVAMEFVNPVFLCWPARLLNADVNNPEAGQYMGNI 352
QY 351 NVTFPSSQTL 351
DB 353 KITFPSSQTL 353

RESULT 3
US-08-483-101-5
; Sequence 5, Application US/08483101
; Patent No. 5932715
; GENERAL INFORMATION:
; APPLICANT: Scott, June R.
; APPLICANT: Froehlich, Barbara
; APPLICANT: Caron, Judy
; TITLE OF INVENTION: CS2 Proteins and Coding Sequences
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee and Winner, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,101
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33878
; REFERENCE/DOCKET NUMBER: 6-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-483-101-5

```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.024;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

QY 1 MNKILFTLFFSSVLFTFAVSADKIPGDSITNIIFGRDRN--ESSPKENILNNEITAY 58
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1 LKKVIVLWFLCSQVYGOSWHTNVEAGSINKTESIGPIDR3AAAS7AHVITHEAVAGY 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 SE8HTLYDRMTFFCLSHNHTLNGACPTSENPPSSSVSGETNITLOFTEKSLIKRELOIK 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 NK0KSLFDRMTFLCMSSTDASKGACPTGSKSS--QGETNIKLIFFTEKSLARKTLNLX 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 GYKOLLPKSYNC---PSGLTINSAHEPNCNKAASGASLYLYIPAGELKNIPLFGGIWDAIL 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 GYKFLYDSCDIHVYKXNLSNTHVVCVSGFTRGVDFTLVPGEDICLLTGGINAAIL 178
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 KLRVKRYSETGYTITITIKLTDKGNIQIWLQPKSDARVDLNLNRPRTGGTVIGNSV 235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 ELVRKHVDYNEGYKYNIITVDLTDKGNIQVWTFKFSIDPRIDLNLNRPNGKSGSNVL 238
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 DMCYDGYSTNSSLLEIRFQDNPKSDGKFLRKINDDTKEIATYLSILLAGKSLITNG 295
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 239 EMCLYDGYSTHSQSIEMRFQDSQTNNEYNLTKTGEPLKPLKLLSLLGGREFFPNNG 298
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 TSLNIAAASLETNWNRITATVMPBEISVPLVLCWPGRLQLDKAVENPAGQYMGNIINVTFT 355
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 299 EAFINDTSSLFNWRILKSVLSPLQISPLVLCWLANLTFMSELNPNRAGEYSGILNVTFT 358
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 PSSOTL 361
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 359 PSSSSL 364
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,932
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.024;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

QY 48 HN-LNNHITAYSSHTLYDRMTFLCSSLSENTLNGACPTSENPPSS-----SVSGETNITL 102
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 816 YNEYSKH--AINSSHNL-----TILGNVTLCG-----ENSSSITCNTNITNKANVTL 862
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 QF-----TEBSLJK3LQIKGYKQLLFKSVNCP5GLTL-----NSAHFNON--- 144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 863 QADTSNNTGLKRTTLFLGNISVEGNLSLTCANANIVCNLSIAEDSTFKGEASDNLNITG 922
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 ---KMAASGASLYLYIPAGELKNIPLFGGIWDATILKLRKRYSETYGT-YTINTITKLT 200
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 923 TFTNNGTANINIKGVKLGIDINK--CG-----LNTTNASGTQKTIINCNTI---NE 970
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 201 KGNITQIWLQPKSDARVDL--NLNRPRTGGTVIGNSVDMCFYDGYSTNSSLLEIRFQ--- 255
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 971 KGDLLN--KNIKADAEQIGNISKKEGKNTLTISSDKVNI-----TNQITIKAGVEGR 1022
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 256 -DNNPKSDGKFLRKINDDTKEIATYLSILLAG--XSLTPNGTSLNIAAASLETNWK 311
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1022 SDSSEAFENANLTIQ-----TKELKLAGDLNLSGKNKABITAKNGSDLTIGNASGNADAK 1076
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 RITATVMPBEISVPLVLCWPG-RLQLDKAVENPAGQYMGNIINVT 353
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1077 K---VTFDKVKDSKITSTGHNVTINSEVKTNGSSNAGNDNST 1116
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-08-362-525-2
; Sequence 2, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTRIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202090.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
```


REGISTRATION NUMBER: 16,773
 REFERENCES/DOCKET NUMBER: 213289/T7020(V)
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 861-3000
 TELEFAX: (202) 822-0944
 TELEX: 5714627 CUSH
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 650 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-362-525-2

Query Match 6.0%; Score 114; DB 3; Length 650;
 Best Local Similarity 20.9%; Pred. No. 0.01;
 Matches 98; Conservative 52; Mismatches 147; Indels 172; Gaps 23;
 QY 4 ILFIPTLFFSSVL-----FTFA-----VSADKTPG-----DESTNIFGPRDRNE--- 43
 DB 8 ILWFLSALASALININDITFSNLEITPLTANKOPDQGTATPDSIADASSIREGDEFTL 67
 QY 44 SSP---KHNILNNHYA-----YSBSHTLYDRMTPLC-----LSSSENTLN 80
 DB 68 SMPHYVRIKILNSSQTATISLADGTEAPKCYVSCQAAYIYENTTFTCTAQNLISSNTLID 127
 QY 81 GACPTSENPSSSVSGETNITLQFTKRSLIKRELOIKGYKQLLFKSVNC-PSGLTINSA 139
 DB 128 GSITFSLNFDGSSVEYEL-----ENAKFKGPMVLKLGQMVDVNFDPAAFTENVF 182
 QY 140 HFNCNKNAASGASLYL--YIPAGELKNLPFGGIWDATLKURVKRYSEYTGVTYINITIK 197
 DB 183 HSGRSTGYGSGFHYLGMCPNGY?-----LGG-----TEKI-----DYDSSNNVLD- 225
 QY 198 LTPKGNIOI-----WLPQKSDARVDLNLRPCTGGGYVIGRN---SVDMCFYDG--- 242
 DB 226 --DCSSVQVYSSNDFNDWPF?QSYNDTNADV-----TCFGSNLWITLDEKLYDGYL 275
 QY 243 -----YSTNSSSLE--IRPDNN--PKSDGKF 265
 DB 276 WYNALQSLPANVTIDHALRFQYTCIDTIANTYVATQFTTRRFIVYQGNLGTASAKSS 335
 QY 266 YLRKINDDTKEI---AVTLS-----LLAGKSLTPPTNGTSLNTADAA 304
 DB 336 FISITITDLTSINTSAYSTGISTSVETGNRTTSVISHVVVTSTKLGTATTSLTIAQTS 395
 QY 305 SLEINWK-----RITAVTMPEISVPVLCWPGRL 332
 DB 396 IYSTDSNITVGTDIHTTSEVISQVETISRETASTVVAAPTSTGTGTGAM 444

RESULT 6

US-09-206-942-41
 ; Sequence 41, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; FILE REFERENCE: Molecular Weight Proteins
 ; CURRENT APPLICATION NUMBER: US/09/206,942
 ; CURRENT FILING DATE: 1998-12-08
 ; EARLIER FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 41
 ; LENGTH: 1005
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 ; US-09-206-942-41

Query Match 5.9%; Score 111.5; DB 4; Length 1005;
 Best Local Similarity 20.3%; Pred. No. 0.037;
 Matches 58; Conservative 65; Mismatches 147; Indels 55; Gaps 14;
 QY 60 ESHTLYDRMTPLCLSSHTNLGACPTSPSSSVSGTETNITLQFTKRS-LIKRELQIK 118
 DB 363 EKNAIFSTHNLTLGGNVTLGGSSSSIKGNINNSKANVTLQAHAGTSHLDKKE--- 418
 QY 119 GYKQLLFKSVNCPGSLTL--NSAHFNCNKNAASGASLYLYIPAGEL-KNLPFGGIW--DA 173
 DB 419 --RTLTLGNVSVGGNLIIGNSNAHIDGNLSIAESAKF-----QCKTNNNLNITGFTTNG 471
 QY 174 TLKLRVKRYSEYTGVTYINITIKLTD-----KGNI-----QIWLPOPKSDARVD 218
 DB 472 TADINIKQGVKLOQDITNNGNLNITNASVNQKTIINGNITNKKGDLNLIKDKANAEIQ 531
 QY 219 L--NLRPCTGGGYVIGRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSD-GKFYLRKINDDTK 275
 DB 532 ICGNISQKEGNTLITSSDKINI-----TKSIELKADTDQGNSDSGVASNANLTIKTK 582
 QY 275 EIAYTLSELLAG---KSLTPNGTSLNITADAASLETNWRNRIITAVTMPEISVPVLCWPGRL 332
 DB 583 ELTLDNLNISGENKAEITAKDNSDLIIGKASSDNSNAKQITFDKVKDSKIS--AGNHNV 640
 QY 333 QLDKAVENPEA-----QYMGNIINVTFTPSQTL 361
 DB 641 TLNSKVETSDSGSTGNGSDNNIGLITSAKDVIV 675

RESULT 7

US-09-206-942-39
 ; Sequence 39, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; FILE REFERENCE: Molecular Weight Proteins
 ; CURRENT APPLICATION NUMBER: JS/09/206,942
 ; CURRENT FILING DATE: 1998-12-08
 ; EARLIER FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 39
 ; LENGTH: 1011
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 ; US-09-206-942-39

Query Match 5.9%; Score 111.5; DB 4; Length 1011;
 Best Local Similarity 20.3%; Pred. No. 0.037;
 Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;
 QY 60 ESHTLYDRMTPLCLSSHTNLGACPTSPSSSVSGTETNITLQFTKRS-LIKRELQIK 118
 DB 369 EKNAIFSTHNLTLGGNVTLGGSSSSIKGNINNSKANVTLQAHAGTSHLDKKE--- 424
 QY 119 GYKQLLFKSVNCPGSLTL--NSAHFNCNKNAASGASLYLYIPAGEL-KNLPFGGIW--DA 173
 DB 425 --RTLTLGNVSVGGNLIIGNSNAHIDGNLSIAESAKF-----QCKTNNNLNITGFTTNG 477
 QY 174 TLKLRVKRYSEYTGVTYINITIKLTD-----KGNI-----QIWLPOPKSDARVD 218
 DB 478 TADINIKQGVKLCQDITNNGNLNITNASVNQKTIINGNITNKKGDLNLIKDKANAEIQ 537
 QY 219 L--NLRPCTGGGYVIGRNSVDMCFYDGYSTNSSSLEIRFQDNNPKSD-GKFYLRKINDDTK 275
 DB 538 ICGNISQKEGNTLITSSDKINI-----TKSIELKADTDQGNSDSGVASNANLTIKTK 588

[illegible]

```

RESULT 10
US-03-242-913B-15
; Sequence 15, Application US/09242913B
; Patent No. 6551811
; GENERAL INFORMATION:
; APPLICANT: FONTAINE, THIERRY
; APPLICANT: HARTLAND, ROBERT
; APPLICANT: MOUINA, ISABELLE
; APPLICANT: LATGE, JEAN-PAUL
; TITLE OF INVENTION: METHOD FOR SORTING ANTIFUNGAL MOLECULES ACTING ON THE
; TITLE OF INVENTION: GLUCANOSYLTRANSFERASE ACTIVITY
; FILE REFERENCE: 05986-0007
; CURRENT APPLICATION NUMBER: US/09/242,913B
; CURRENT FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: PCT/FR97/01540
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 60/024,910
; PRIOR FILING DATE: 1996-08-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-03-242-913B-15

```

RESULT 11

US-09-206-942-45
 ; Sequence 45, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel E.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; FILE REFERENCE: 1038-861 M18:jb
 ; CURRENT APPLICATION NUMBER: US/09/206,942
 ; CURRENT FILING DATE: 1998-12-08
 ; EARLIER APPLICATION NUMBER: 09/167,568
 ; EARLIER FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 45
 ; LENGTH: 1095
 ; TYPE: PRT
 ; ORGANISM: Haemophilus influenzae
 ; US-09-206-942-45

RESULT 12
 US-09-206-942-43
 ; Sequence 43, Application US/09206942
 ; Patent No. 6432669
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
 ; TITLE OF INVENTION: Molecular Weight P-O-r-e-i-n-s
 ; FILE REFERENCE: 1038-861 MIS:jb
 ; CURRENT APPLICATION NUMBER: US/09/206,942
 ; CURRENT FILING DATE: 1998-12-08
 ; EARLIER APPLICATION NUMBER: 05/167,568
 ; EARLIER FILING DATE: 1998-10-07
 ; NUMBER OF SEQ ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 43
 ; LENGTH: 1101
 ; TYPE: prt
 ; ORGANISM: Haemophilus influenzae
 US-09-206-942-43

Db 489 GEVKSAGNVNITASGNTLVNSITQNVTVTAN-----SGAITTEGST----- 533
QY 332 LQDAKVENPEAGQYMGNN-----VTFTPSQTL 361
Db 534 --INATTGDANITQTGNGINGKVESSGSVTLIATGQTL 570

RESULT 15
US-09-206-942-37
; Sequence 37, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: Molecular Weight Proteins
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-37

Query Match 5.4%; Score 101; DB 4; Length 1222;
Best Local Similarity 21.1%; Pred. No. 0.61;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;
QY 21 VSADKIPGESITNIFGPR----DRNESSPKH-----NILN-----NHITAYS 59
Db 559 ISSDKVNITERITIKAGVNGDSDNEATSANLTIKELKLTNDLNI SGFNKAEITAKD 618
QY 60 ESH-TLYD-----RMTR-----LCLSSHN-TLNGACPTSENPSSSSVSGETNI 100
Db 619 NSKLTIGDSDAGNTDAKVTFSNVKDSKISASDHNVTLNSKVETSGDITDSTEDGGNNNT 678
QY 101 TLQFTKRSLIKRELIQKGYQL---LFKSVNGCPSGLTLSAHFNCKNAASGASLYLYI 157
Db 679 GLTIITAKNVTVNN--NITSEKTVNITASENVTTKAGTTIN-----ATTGSVEVTA 726
QY 158 PAGELKNLPFGGIWDATLKURVKRKRSEYTGTYTINIKITDKGNLIQIWLPOFKSDARV 217
Db 727 KTGDIK-----GGI-----ESNSGNVNI----- 744
QY 218 DLNLRPTGGTYI-----GENSVDMCFYDGYSTNSSLSLEIRFQDNNPKSDGKPYLFRKIN 271
Db 745 -----TASGDTLVNSNITQCN-VTVAAASGAVTTKGTSTINATTGNANITTK--TGEIN 795
QY 272 DDTKEIAYTLLAGKSLTPTNGTSLNIADASLETNWNRIATVMPREISVPVLCWPGR 331
Db 796 GEVKSAGNVNITASGNTLVNSITQNVTVTAN-----SGAITTEGST----- 840
QY 332 LQDAKVENPEAGQYMGNN-----VTFTPSQTL 361
Db 841 --INATTGDANITQTGNGINGKVESSGSVTLIATGQTL 877

Search completed: May 6, 2004, 10:20:19
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 10:19:17 ; Search time 48 Seconds
(without alignments)
2087.533 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFIPTLFSSVLTFA.....BAGQVMGNVTFPSSOTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277565755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA.*
- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/prodata/1/pubpaa/CT_NEW_PUB.pep.*
 - 3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/prodata/1/pubpaa/CTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	100.0	361	9	US-09-839-894-10
2	1862.5	98.8	361	9	US-09-839-894-28
3	1542	81.8	359	9	US-09-839-894-29
4	937	49.7	362	9	US-09-839-894-31
5	916.5	48.6	364	9	US-09-839-894-30
6	247.5	13.1	353	9	US-09-839-894-32
7	120	6.4	1335	12	US-10-282-122A-42853
8	119.5	6.3	1770	9	US-09-303-232-2
9	116.5	6.2	1600	13	US-10-092-880-10
10	114	6.0	650	15	US-10-369-493-22177
11	113.5	6.0	513	12	US-10-282-122A-61451
12	111.5	5.9	1005	14	US-10-193-764-4
13	111.5	5.9	1021	14	US-10-193-764-39
14	109	5.8	2383	9	US-09-912-020-302
15	109	5.8	2383	12	US-10-282-122A-42852

16	108	5.7	524	14	US-10-347-278-15	Sequence 15, Appl
17	108	5.7	524	14	US-10-347-252-15	Sequence 15, Appl
18	108	5.7	2529	12	US-10-282-122A-58834	Sequence 58834, A
19	105	5.6	598	12	US-10-282-122A-60611	Sequence 60611, A
20	103.5	5.5	692	15	US-10-369-493-23473	Sequence 23473, A
21	103.5	5.5	2399	12	US-10-282-122A-59054	Sequence 59054, A
22	101	5.4	915	14	US-10-193-764-35	Sequence 35, Appl
23	101	5.4	1222	14	US-10-193-764-37	Sequence 37, Appl
24	101	5.4	1228	14	US-10-193-764-34	Sequence 34, Appl
25	101	5.4	1599	13	US-10-092-880-9	Sequence 9, Appl
26	100.5	5.3	1004	14	US-10-193-764-53	Sequence 53, Appl
27	100.5	5.3	1010	14	US-10-193-764-51	Sequence 51, Appl
28	99.5	5.3	969	14	US-10-193-764-32	Sequence 32, Appl
29	99.5	5.3	975	14	US-10-193-764-30	Sequence 30, Appl
30	99.5	5.3	1300	12	US-10-282-122A-67412	Sequence 67412, A
31	99	5.2	839	10	US-09-738-269-23	Sequence 23, Appl
32	99	5.2	839	13	US-10-023-437-23	Sequence 23, Appl
33	99	5.2	992	14	US-10-193-764-57	Sequence 57, Appl
34	99	5.2	998	14	US-10-193-764-55	Sequence 55, Appl
35	99	5.2	1848	10	US-09-839-996-6	Sequence 6, Appl
36	99	5.2	1848	12	US-10-645-655-6	Sequence 6, Appl
37	99	5.2	1848	14	US-10-080-505-6	Sequence 43, Appl
38	97	5.1	977	14	US-10-193-764-49	Sequence 47, Appl
39	97	5.1	627	14	US-10-193-764-47	Sequence 47, Appl
40	96.5	5.1	627	12	US-10-425-114-62687	Sequence 62687, A
41	95.5	5.1	925	14	US-10-101-464A-922	Sequence 922, App
42	95.5	5.1	1194	12	US-10-282-122A-46577	Sequence 46577, A
43	95	5.0	867	9	US-09-839-894-6	Sequence 6, Appl
44	94.5	5.0	312	12	US-10-424-599-234245	Sequence 234245
45	94.5	5.0	2353	9	US-09-797-862-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1
US-09-839-894-10
; Sequence 10, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altobum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; TITLE OF INVENTION: CSA OPERON
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 361
; TYPE: PRT
; ORGANISM: E. coli
US-09-839-894-10

Query Match	100.0%;	Score	1886;	DB	9;	Length	361;
Best Local Similarity	100.0%;	Pred. No.	3.3e-181;				
Matches	361;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
Qy	1	MNKLFIPTLFSSVLTFAVSADKTPGDSINIFGPRDRNSSPKNNHITAYSE	60				
Db	1	MNKLFIPTLFSSVLTFAVSADKTPGDSINIFGPRDRNSSPKNNHITAYSE	60				
Qy	61	SHLYDRMTLCSSHNTLNAGCTSPNPPSSSVSGTNTLTQTEKRSIKRELOIKGY	120				
Db	61	SHLYDRMTLCSSHNTLNAGCTSPNPPSSSVSGTNTLTQTEKRSIKRELOIKGY	120				
Qy	121	KQLIFSVNCPGSLTNSAHFNQNAASGASLYIYPAGELKNLPFGGTWDTKLKRVK	180				

Db 121 KOLLFKSVNCPGSLTNSAHFNCNNAAGSALYIYPAGELKNIPFGGIWDATLKLVRK 180
QY 181 RYSETYGTITNITIKLTDKGNIQIWLPOFKSDARVDNLPTGGGTIGRNSVDMCFY 240
Db 181 RYSETYGTITNITIKLTDKGNIQIWLPOFKSDARVDNLPTGGGTIGRNSVDMCFY 240
QY 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLSLLAGKSLTPNGTSLNI 300
Db 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLSLLAGKSLTPNGTSLNI 300
QY 301 ADAASLETNNRITAVTMPISVPLVLCWPGRLQIDAKVENPEAGQYMGNIINVTFTSSQT 360
Db 301 ADAASLETNNRITAVTMPISVPLVLCWPGRLQIDAKVENPEAGQYMGNIINVTFTSSQT 360
QY 361 L 361
Db 361 L 361

RESULT 2

US-09-839-894-28
; Sequence 28, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-28

Query Match 98.8%; Score 1862.5; DB 9; Length 361;
Best Local Similarity 99.4%; Pred. No. 7.8e-179;
Matches 359; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE 60
Db 2 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE 61
QY 61 SHTLYDRMTFLCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 120
Db 62 SHTLYDRMTFLCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 121
QY 121 KOLLFKSVNCPGSLTNSAHFNCNNAAGSALYIYPAGELKNIPFGGIWDATLKLVRK 180
Db 122 KOLLFKSVNCPGSLTNSAHFNCNNAAGSALYIYPAGELKNIPFGGIWDATLKLVRK 181
QY 181 RYSETYGTITNITIKLTDKGNIQIWLPOFKSDARVDNLPTGGGTIGRNSVDMCFY 240
Db 182 RYSETYGTITNITIKLTDKGNIQIWLPOFKSDARVDNLPTGGGTIGRNSVDMCFY 241
QY 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLSLLAGKSLTPNGTSLNI 300
Db 242 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLSLLAGKSLTPNGTSLNI 300
QY 301 ADAASLETNNRITAVTMPISVPLVLCWPGRLQIDAKVENPEAGQYMGNIINVTFTSSQT 360
Db 301 ADAASLETNNRITAVTMPISVPLVLCWPGRLQIDAKVENPEAGQYMGNIINVTFTSSQT 360

QY 361 L 361
Db 361 L 361

RESULT 3

US-09-839-894-29
; Sequence 29, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
US-09-839-894-29

Query Match 81.8%; Score 1542; DB 9; Length 359;
Best Local Similarity 80.9%; Pred. No. 1.6e-146;
Matches 292; Conservative 32; Mismatches 35; Indels 2; Gaps 2;

QY 1 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE 60
Db 1 MNKILFIPTLFFSSVLTFAVSADKIPGDESITNIFGPRDRNESSPKHNLNNHITAYSE 60
QY 61 SHTLYDRMTFLCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 120
Db 61 SHTLYDRMTFLCLSHNTLNACPTSENPPSSSVSGETNITLQFTEKSLIKRELQIKGY 120
QY 121 KOLLFKSVNCPGSLTNSAHFNCNNAAGSALYIYPAGELKNIPFGGIWDATLKLVRK 180
Db 121 KOLLFKSVNCPGSLTNSAHFNCNNAAGSALYIYPAGELKNIPFGGIWDATLKLVRK 180
QY 181 RYSETYGTITNITIKLTDKGNIQIWLPOFKSDARVDNLPTGGGTIGRNSVDMCFY 240
Db 181 RYSETYGTITNITIKLTDKGNIQIWLPOFKSDARVDNLPTGGGTIGRNSVDMCFY 239
QY 241 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLSLLAGKSLTPNGTSLNI 300
Db 240 DGYSTNSSSLEIRFQDNPKSDGKFLRKINDDTKEIAYTSLSLLAGKSLTPNGTSLNI 299
QY 301 ADAASLETNNRITAVTMPISVPLVLCWPGRLQIDAKVENPEAGQYMGNIINVTFTSSQT 360
Db 300 -NTASLETNNRITAVTMPISVPLVLCWPGRLQIDAKVENPEAGQYMGNIINVTFTSSQT 358
QY 361 L 361
Db 359 L 359

RESULT 4

US-09-839-894-31
; Sequence 31, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland

1 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
2 TITLE OF INVENTION: CSA OPERON
3 FILE REFERENCE: UOFMD.006A
4 CURRENT APPLICATION NUMBER: US/09/839,894
5 CURRENT FILING DATE: 2001-04-20
6 PRIOR APPLICATION NUMBER: 60/198,626
7 PRIOR FILING DATE: 2000-04-20
8 NUMBER OF SEQ ID NOS: 40
9 SOFTWARE: FastSeq for Windows Version 4.0
10 SEQ ID NO 31
11 LENGTH: 362
12 TYPE: PR1
13 ORGANISM: Artificial Sequence
14 FEATURE:
15 OTHER INFORMATION: ETEC Protein Homology Sequence
16 US-09-839-894-31

Query Match 49.7%; Score 937; DB 9; Length 362;
Best Local Similarity 53.1%; Pred. No. 1.7e-85;
Matches 197; Conservative 54; Mismatches 98; Indels 22; Gaps 12;
1 3 KILFIPTLPSSVLFTFAVSADKIPGDE--SIITNIF-GER-DRNESSPKHNLNNHITAY 58
2 2 KXIFPL----SIIFSAVVSAGRYPETTVGNLTKSQAPRLDSVQSPINFTNVAGY 57
3 59 SPSHTLYDRMTFLCLSSHTNLGACPTSENPSSSVS-GETNITLQFTKRSLIKELQI 117
4 58 SLASHSLYDRAIVFLCTSSNPVNGACPTI---GTSGVQYGTITLQFTKRSLIKENINI 114
5 118 KGVKQLLFKSVNCPG--LTINSAHFNCNKNA--ASGASLYLYIPAGELKNLPFGGWDAT 174
6 115 AGNKKPIWENQSCDFSNIMVINSKSCGANGTILNMLYIPAGEINKLPFGGWEAT 174
7 175 LKLRVKRRYSET----YGIYTNITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGGTII 230
8 175 LILRLS-RGEVSSSTHYGNTVNTLVDLTDKGNIQWLPGFHSNPRVDLNRPIGNKYIS 233
9 231 GRNSVDMCFYDGYSTNSSLLEIRPDNNPKSDGKFYLKINDTKETAYTLSSLACKSL 290
10 234 GSNLDMCFYDGYSTNSDSMVIRKFDQDNPNTSSEYNLYKIG-GTEKLPYAVS-LIGBKIF 291
11 291 TPNGTSLNLTADAASLETNNRITAVTNPISVPLVCPGRLQIDAKVENPEAGQYMGNI 350
12 292 YPVNGQSFINDSSVLETNWNRVTAVAMPEVNVPLVCPGRLQIDAKVENPEAGQYSGQI 351
13 351 NVFTPTSSQTL 361
14 352 YIIFTPSEVNL 362

RESULT 5
US-09-839-894-30
; Sequence 30, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 364
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
; US-09-839-894-31

1 OTHER INFORMATION: ETEC Protein Homology Sequence
2 US-09-839-894-30

Query Match 48.6%; Score 916.5; DB 9; Length 364;
Best Local Similarity 48.1%; Pred. No. 2e-83;
Matches 176; Conservative 63; Mismatches 120; Indels 7; Gaps 3;

1 1 MKKILFIPTLPSSVLFTFAVSADKIPGDESIITNIFGPDNRN--BSSPKHNLNNHITAY 58
2 1 MKKVIFVLSMFLCSQVYQSWHTNVEAGSINTFSGIDRSAASYPDAHYFHFREVAGY 60
3 59 SPSHTLYDRMTFLCLSSHTNLGACPTSENPSSSVS-GETNITLQFTKRSLIKELQI 118
4 61 NKDHSLSFDRMTFLCMSSDASKGACPTGNSKSS--QGETNIKLIPTKRSKLARKTLNLK 118
5 119 GYKQLLFKSVNCPG--PSGLTINSAHFNCNKNAASGASLYLYIPAGELKNLPFGGIWDAT 175
6 119 GYKRFLYESDRCIHYVDKMLNLSHTVCKVGSFTRGVDFLYIPQGHIDGLLTGGIWKATL 178
7 175 KLRVKRRYSETYCTYTNITIKLTDKGNIQIWLPOFKSDARVDLNRPTGGGTIYGENSV 235
8 179 ELRVXRHYDYNHGTYKVNITVDLTKGNIQVWTPKPHSPRIDLNRPGNGKYSGSNVL 238
9 236 DMCFYDGYSTNSSLLEIRPDNNPKSDGKFYLKINDTKETAYTLSSLACKSLPTMG 295
10 239 EMCLYDGYSTHSQSIEMR--QDDSQTCNNEVNLKTGEPLKGLPYK-SILLGGREFYPNNG 298
11 296 TSLNTADAASLETNNRITAVTNPISVPLVCPGRLQIDAKVENPEAGQYMGNIYVTF 355
12 299 KAPTINDTSSLFINNKRKISVLPQISIPVLCWPAANLTFMSELNPEAGEYSGLNVTF 358
13 356 PSSQTL 361
14 359 PSSSSL 364

RESULT 6
US-09-839-894-32
; Sequence 32, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Altboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: UOFMD.006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 353
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ETEC Protein Homology Sequence
; US-09-839-894-32

Query Match 13.1%; Score 247.5; DB 9; Length 353;
Best Local Similarity 28.9%; Pred. No. 5.5e-16;
Matches 89; Conservative 42; Mismatches 126; Indels 51; Gaps 13;

1 77 NTANG-ACPTSENPSSSS-----VSGETNITLQFTKRSLIKELQIKGYKQLLFK 126
2 72 NVLGGWVCRNRNNEGCEETHLVWYAFGAYSILRFRQISHAEITL----- 120
3 127 SVNCPGSLTINSAHFNC-----NKNAAAS---GASLYLYIPAGELKNLPFGGIWDATLKL-R 178
4 121 -----ILLGSVRDACTGVINWNAACQWGRSLKLRIPSELAKIPTSGTWKATLVLDY 173

[illegible]

RESULT 7

```

US-10-282-122A-42853
? Sequence 42853, Application US/10282122A
? Publication No. US20040029129A1
?
? GENERAL INFORMATION:
? APPLICANT: Wang, Liangsu
? APPLICANT: Zamudio, Carlos
? APPLICANT: Malone, Cheryl
? APPLICANT: Haselbeck, Robert
? APPLICANT: Orlsen, Kari
? APPLICANT: Zyskind, Judith
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John
? APPLICANT: Carr, Grant
? APPLICANT: Yamamoto, Robert
? APPLICANT: Forsyth, R.
? APPLICANT: Xu, H.
?
? TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
?
? FILE REFERENCE: EMTA.034A
?
? CURRENT APPLICATION NUMBER: US/10/282,122A
?
? CURRENT FILING DATE: 2003-02-20
?
? PRIOR APPLICATION NUMBER: 60/131,078
?
? PRIOR FILING DATE: 2000-03-21
?
? PRIOR APPLICATION NUMBER: 60/206,848
?
? PRIOR FILING DATE: 2000-05-23
?
? PRIOR APPLICATION NUMBER: 60/207,727
?
? PRIOR FILING DATE: 2000-05-26
?
? PRIOR APPLICATION NUMBER: 60/230,335
?
? PRIOR FILING DATE: 2000-09-06
?
? PRIOR APPLICATION NUMBER: 60/230,347
?
? PRIOR FILING DATE: 2000-09-09
?
? PRIOR APPLICATION NUMBER: 60/242,578
?
? PRIOR FILING DATE: 2000-10-23
?
? PRIOR APPLICATION NUMBER: 60/253,525
?
? PRIOR FILING DATE: 2000-11-27
?
? PRIOR APPLICATION NUMBER: 60/257,931
?
? PRIOR FILING DATE: 2000-12-22
?
? PRIOR APPLICATION NUMBER: 60/267,636
?
? PRIOR FILING DATE: 2001-02-09
?
? PRIOR APPLICATION NUMBER: 60/269,308
?
? PRIOR FILING DATE: 2001-02-16
?
? Remaining Prior Application data removed - See File Wrapper or PALM.
?
? NUMBER OF SEQ ID NOS: 78614
?
? SOFTWARE: PatentIn version 3.1
?
? SEQ ID NO 42853
?
? LENGTH: 1335
?
? TYPE: PRT
?
? ORGANISM: Escherichia coli
?
US-10-282-122A-42853

```

[illegible]

RESULT 8

```

US-09-303-232-2
; Sequence 2, Application US/09303232A
; Patent No. US20020036657A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; TITLE OF INVENTION: Nucleic Acids which encode
;             insect acetylcholine receptor subunits
; FILE REFERENCE: Je A 33 020-Foreign Countries
; CURRENT APPLICATION NUMBER: US/09/303,232A
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: DE 198 19 829.9
; EARLIER FILING DATE: 1998-05-04
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 770
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-303-232-2

```

Query Match	5.3%;	Score 119.5;	DB 9;	Length 770;
Best Local Similarity	20.4%;	Pred. No. 0.015;		
Matches 83;	Conservative	48;	Mismatches 140;	Indels 135; Gaps 20;
QY	9	TLRFSSVLFPAYSAKDIPGDESITNIFGPRDRNESSPKFINLNH-----	54	
DB	249	TIATISYLGSPAAGLKHSSSSSSSN-----SSNSSTQILNGLNKHSWIFLLIYLNLSAK	304	
QY	55	--TAYASEHTLYDRMTFCLSSHNTLNGACPTSENFSSSSVSSETNITLOFTEKRSLIK	112	
DB	305	VCLAGVHEKRLH-----LIDPYNTILERPVLNESDPLQLSFGTLTMOITDIDVBNKOLLY	359	
QY	113	RELQIKGYKQLLPKSVNCPGSLTNSAHFNCKNAAGSALYLYPAGELKNLPFGGIWD	172	
DB	360	TNVWLK-----LEWMDNLRWNTSDYGVK-----DLR-IP2HRIWK	395	
QY	173	ATLKLRLVRRYSEF-----YGYHTYNTIITKTDKNIQIWIPO-----FKSDARVDLNLRTGG	226	
DB	396	PDVFM-----YNSADEGFDGTQYNVVR-----NNGSC-LYVPPGFIKCTCKIDITWFP----	444	
QY	227	GTIYGRNSVDMCF-----YDGYSTNSSSLETRFQDNNPKS-----DGKFLYLRKINDDTKE	276	


```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 1005
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-10-193-764-41

Query Match      5.9%; Score 111.5; DB 14; Length 1005;
Best Local Similarity 20.3%; Pred. No. 0.14; Indels 55; Gaps 14;
Matches 68; Conservative 65; Mismatches 147;

QY 60 ESHTLYDRMTFLCISSHNTLNGACPTSENPPSSSVSGETNITLQFTKRS-LIKRELQIK 118
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 363 EKNAIPSTHNLTLGCVTLGSESSSSNKGNNINNSKANVTLQAHAGTSHLDKKE---- 418
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 GYKQLLPKSVNCPGSLTL--NSAHFNCKNAASGASLYLYIPAGEL-KNLPFGGIW--DA 173
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 419 --RTLTLGNVSVGGNLTIGSNAHIDGNLSIAESAKF-----QGTNNNLNLTGFTNG 471
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 TLKLRVKRYSYTYGTITNITIKLTD-----KGNI-----QIWLQPKSDARVD 218
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 TADINIKQGVKLDGDTTNGNLNITNASVNQKTIINGNITNKKGDNLKDIKANAETQ 531
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 219 L--NLRETGGTYIGNSVDMCFYDGYSTNSSSLEIRPQDNNPKSP-GKPYLRKINDDTK 275
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 532 IGNISQKEGNLTISDDKINI-----TKRIBIKADTDGNSDSGVASNAALTIKTQ 582
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 276 EIAYTLSELLAG---KSLPTNGTSLNIADAASLETNWRITAVTPEISVPVLCWPGEL 332
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 583 ELTLDNLNLSGPNKAEITAKDNLIIIGKASSDLSNAKQITPDKVKDKIS--AGHNHV 640
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 333 QLDKAVENPEA-----GQVMGNINVTFTPSSQTL 361
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 641 TLNSKVETNSDGSSTGNGSDNNIGLTIISAKVTV 675
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-10-193-764-39
; Sequence 39, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michael H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239M-S
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 1011
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-10-193-764-39

Query Match      5.9%; Score 111.5; DB 14; Length 1011;
Best Local Similarity 20.3%; Pred. No. 0.14; Indels 55; Gaps 14;
Matches 68; Conservative 65; Mismatches 147;

QY 60 ESHTLYDRMTFLCISSENTLNGACPTSENPPSSSVSGETNITLQFTKRS-LIKRELQIK 118
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 EKNAIFSTHNLTLGCVTLGSESSSSNKGNNINNSKANVTLQAHAGTSHLDKKE---- 424
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 119 GYKQLLPKSVNCPGSLTL--NSAHFNCKNAASGASLYLYIPAGEL-KNLPFGGIW--DA 173
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 --RTLTLGNVSVGGNLTIGSNAHIDGNLSIAESAKF-----QGTNNNLNLTGFTNG 477
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 TLKLRVKRYSYTYGTITNITIKLTD-----KGNI-----QIWLQPKSDARVD 218
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61451
; LENGTH: 513
; TYPE: PRF
; ORGANISM: Legionella pneumophila
US-10-282-122A-61451

Query Match      6.0%; Score 113.5; DB 12; Length 513;
Best Local Similarity 21.7%; Pred. No. 0.032; Indels 107; Gaps 16;
Matches 76; Conservative 51; Mismatches 116;

QY 50 ILNNHITAVSEHTLYDRMTF---LCLSSHTLNGACPTSENPPSSSVSGETNITLQFT 105
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 LNLKIPPLKGFHVFSEINIDGLKLEINQEN-----NEKPLSEKNAQNYNITYP 141
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 106 EKESLIKRE-QIKGYKQLLPKSVNCPGSLTLNSAHFNCKNAASGASLYLYIPAGELKNL 165
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 NEQFAICKLLSHG-QIILNS-----KGHSTVLKNIQIGARQF-----NLKNS 183
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 166 PPGGIWDTLKLVRKRYSETYGTITNITIKLTDKNIQIWLQPKSDARVDNLKPTG 225
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 PF-----SVQIKAKLTDAA-----FLOTAKANINFK--- 209
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 226 GGTYIGNSVDMCFYDGYST--NSSLEIRPQDNNPKSPGKPYLRKINDDTK-----EIAY 279
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 210 ----GRVLSPSIIDEINLSGKSSIEGQLQ-QNIIIN-QPAIKKINTLTKHKRDIQF 263
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 280 ---TLSL-----LLAGKSLPTNG-----TSLNIADAASLETNWRITA 315
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 264 NPLTILSYNGSIGDMQVIATQQLINQATNLDGKQLTISLLKHPAISGNLDYSIHAS 323
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 VTMPEISVFLCWPGRIQL-DAKVENPEAGQVMGNIV-----TFTPSS 358
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 IPLKALSIESLSVSGTITLKDGEVYVNLQDQLLNLKVKLNLSLMTETPDN 373
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-10-193-764-41
; Sequence 41, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michael H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239M-S
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 91

```

Tue May 11 10:31:28 2004

us-09-839-894-10.rapb

```
QY 478 TADINIKQGVVKLOGDITNGNINITTASVNOQTIIINGNITNKGLNLIKIDIKANAEIQ 537
Db 219 L--NLRPPTGGTYIGNSVDMCFYDGYSTNSSLSEIRFQDNNPKSD-GKPYLRKINDDTK 275
Db 538 CGNISQKEGNTISSDKINI-----TKRIBIKADTDQNSDSGVASNANLITIKT 588
QY 276 EYATYLSLLAG---KSLTPTNGTSLINIDASLETNWNRITAVTWPESVPLVCPGRL 332
Db 589 ELTLDNLNISGFNAEITAKGNSDLIIIGKASDSNAKQITFDKVKOSKIS--AGNHV 646
QY 333 QLDKAVENPEA-----GQYMGINKVTFPPSSQTL 361
Db 647 TUNSKVETSNSGSGTNGSGDDNNGIGTISAKOVTV 681

RESULT 14
US-09-912-020-302
; Sequence 302, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ESCHERICHIA COLI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 302
; LENGTH: 2383
; TYPE: PRF
; ORGANISM: E. Coli
US-09-912-020-302

Query Match 5.8%; Score 109; DB 9; Length 2383;
Best Local Similarity 20.0%; Pred. No. 0.97;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLFTFAVSAADKIPGDESITNIGPRDRNESSPKHILNNHITAY---SESHLYDRMT 69
Db 762 SAKIATLSASNGVLANENAATVSVNVADSGS---NPINDHTVTFVAVLSGSAFSFNQN 813
QY 70 FLCLSHNTLNGACPTSENPSSSVSGSTNITLOFTKRSI-----K 112
Db 819 ----TAKTDVNGLA-TFLKSKQEDNTVETVLENGVKQTLIVSFVGSSTAQVDLQSK 873
QY 113 RELQIKGYKQL-LFKSVNCPGSLTILNSAHFNCKNNAAGSGLYLYIPAGELKNIPLFGGIW 171
Db 874 NEVVADGNDSVTMTATVADAKGNLNDVWVTFNVNSABAKLSQTEVNSHD-----GIA 926
QY 172 DATLKLVRKRYSETYGYTYTINIKLTDKNIQIWLQPKSDARVDNLNRPSTGGTYIG 231
Db 927 TATLT-----SLKNGDYRVTSYSSGSQANQQVNFITGDOSTAALTLSV-PSGDIIV-- 976
QY 232 RNSVDMCFYDGYSTNSS---SLERFDNNPKSDGKTVLAKINDDTKEIATYLSLLAG 287
Db 977 -----TNTAPQYMTATLQDKNGNPLKKEITFSVPND-----VASKFSISNGG 1019
QY 288 KSLTPTNGTSL-----NIADAASLETNWNRITAVTWPESVPLV 327
Db 1020 KGMTDSNGVAIASLTGLTAGTHMIVARLANSVSDAQPMFTVADKDRAVVLIQTSKABII 1079
```

```
QY 328 WPG--RLQLDAKVENPEAGQYMGININVTPT 355
Db 1080 GNGVDETLTATVKDP-SNHPVAGITVNT 1108

RESULT 15
US-10-282-122A-42852
; Sequence 42852, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42852
; LENGTH: 2383
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-282-122A-42852

Query Match 5.8%; Score 109; DB 12; Length 2383;
Best Local Similarity 20.0%; Pred. No. 0.97;
Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

QY 13 SSVLFTFAVSAADKIPGDESITNIGPRDRNESSPKHILNNHITAY---SESHLYDRMT 69
Db 752 SAKIATLSASNGVLANENAATVSVNVADSGS---NPINDHTVTFVAVLSGSAFSFNQN 816
QY 70 FLCLSHNTLNGACPTSENPSSSVSGSTNITLOFTKRSI-----K 112
Db 819 ----TAKTDVNGLA-TFLKSKQEDNTVETVLENGVKQTLIVSFVGSSTAQVDLQSK 873
QY 113 RELQIKGYKQL-LFKSVNCPGSLTILNSAHFNCKNNAAGSGLYLYIPAGELKNIPLFGGIW 171
Db 874 NEVVADGNDSVTMTATVADAKGNLNDVWVTFNVNSABAKLSQTEVNSHD-----GIA 926
QY 172 DATLKLVRKRYSETYGYTYTINIKLTDKNIQIWLQPKSDARVDNLNRPSTGGTYIG 231
```

Tue May 11 10:31:28 2004

Db 927 TATLT-----SLKNGDYRVTA SVSSGQANQQVNF IGDSQSTAALTLSV-PSGDITV-- 976
QY 232 RNSVDYCFYDGYSTNS- ---SLEIRFQDNPKSDGKFYLRKINDTKETAYTSLILLAG 287
Db 977 -----TNTAPQYMTATLQDRKGNPLKDKETITFSPND-----VASKFESISNGG 1019
QY 288 KSLTPTNGTSL-----NIADAASLSTNNNRITAVTMPEISVPVLC 327
Db 1020 KGXTDSNGVAIASLTGTLAGTHIMARLANSVSDAQPMTFVADKDRVVVLQTSKAEII 1079
QY 328 WPG--RLQLDKVENPEAGQVMGNINVTFT 355
Db 1080 GNGVDETTLTATVKDP-SNHPVAGITVNET 1108

Search completed: May 6, 2004, 10:25:06
Job time : 49 secs